Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 10 (2008), Issue 1 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Original Paper

Design and Implementation of an Interactive Website to Support Long-Term Maintenance of Weight Loss

Victor J Stevens[1], PhD; Kristine L Funk[1], MS, RD; Phillip J Brantley[2], PhD; Thomas P Erlinger[3], MD, MPH; Valerie H Myers[2], PhD; Catherine M Champagne[2], PhD; Alan Bauck[1]; Carmen D Samuel-Hodge[4], PhD; Jack F Hollis[1], PhD

[1]Kaiser Permanente, Center for Health Research, Portland, OR, USA

[2]Pennington Biomedical Research Center, Baton Rouge, LA, USA

[3]Johns Hopkins University School of Medicine, Baltimore, MD, USA

[4]Duke University Medical Center, Durham, NC, USA

Corresponding Author:
Kristine L Funk, MS, RD
Center for Health Research
Kaiser Permanente
3800 N Interstate Ave
Portland, OR 97227
USA
Phone: +1 503 335 2443
Fax: +1 503 3352424
Email: Kristine.funk@kpchr.org

Abstract

Background: For most individuals, long-term maintenance of weight loss requires long-term, supportive intervention. Internet-based weight loss maintenance programs offer considerable potential for meeting this need. Careful design processes are required to maximize adherence and minimize attrition.

Objective: This paper describes the development, implementation and use of a Web-based intervention program designed to help those who have recently lost weight sustain their weight loss over 1 year.

Methods: The weight loss maintenance website was developed over a 1-year period by an interdisciplinary team of public health researchers, behavior change intervention experts, applications developers, and interface designers. Key interactive features of the final site include social support, self-monitoring, written guidelines for diet and physical activity, links to appropriate websites, supportive tools for behavior change, check-in accountability, tailored reinforcement messages, and problem solving and relapse prevention training. The weight loss maintenance program included a reminder system (automated email and telephone messages) that prompted participants to return to the website if they missed their check-in date. If there was no log-in response to the email and telephone automated prompts, a staff member called the participant. We tracked the proportion of participants with at least one log-in per month, and analyzed log-ins as a result of automated prompts.

Results: The mean age of the 348 participants enrolled in an ongoing randomized trial and assigned to use the website was 56 years; 63% were female, and 38% were African American. While weight loss data will not be available until mid-2008, website use remained high during the first year with over 80% of the participants still using the website during month 12. During the first 52 weeks, participants averaged 35 weeks with at least one log-in. Email and telephone prompts appear to be very effective at helping participants sustain ongoing website use.

Conclusions: Developing interactive websites is expensive, complex, and time consuming. We found that extensive paper prototyping well in advance of programming and a versatile product manager who could work with project staff at all levels of detail were essential to keeping the development process efficient.

Trial Registration: clinicaltrials.gov NCT00054925

(J Med Internet Res 2008;10(1):e1) doi:10.2196/jmir.931

KEYWORDS

Internet; website design; behavioral interventions; weight loss; weight maintenance

Introduction

Obesity has become a major public health problem in the United States [1] with 65% of US adults now overweight or obese [2-4]. Obesity has been linked to increased overall mortality [5-7], decreased life expectancy [8,9], and greatly increased medical care costs [10-12]. The annual US medical expenditure attributable to obesity is estimated to be US $75 billion [13].

National recommendations for weight loss treatment call for intervention programs combining reduced energy intake, improved dietary choices, increased physical activity, and behavior therapy [14]. The most effective format for initial treatment is a series of weekly, professionally led group sessions [15-17]. Longer treatment programs result in greater weight loss, and many weight loss programs now continue initial treatment for 6 months. Immediate health benefits of weight loss include reduced blood pressure and improved blood glucose levels. Sustained reduction of even moderate amounts of weight (4 kg or more) has been shown to significantly reduce the risk of developing hypertension [18,19] and diabetes [20,21] over 3 years.

Even with successful weight loss during the first 6 months of treatment, there is a strong tendency toward weight regain following treatment termination. Although continuing weekly meetings as long as 40 weeks has been shown to be effective in preventing weight regain [17,22], a life-long series of weekly group meetings is not an attractive or practical option. To deal with this problem, there has been considerable interest in developing less intensive, but equally effective, long-term maintenance programs.

Recent reviews suggest that initial weight loss treatment may require different behavioral approaches than weight loss maintenance [15,17,23,24]. Specifically, building calorie counting skills and learning how to select less-energy-dense foods may be more critical for weight loss, whereas use of relapse prevention techniques, problem solving, and enhancing participant motivation may be more germane for weight maintenance. Due to their relatively low cost per person and flexibility of access, alternative communication technologies (eg, Internet) may provide attractive new channels for maintenance interventions [25-28]. For example, some studies comparing weight loss between an Internet-based intervention group and a therapist-led group found that weight loss was similar in the two groups [29]. Recent studies have shown that Internet-based weight loss interventions may be particularly cost-effective when trying to reach a large population [30].

The Weight Loss Maintenance Trial (WLM) (Trial Registration: clinicaltrials.gov NCT00054925) was designed to systematically study the efficacy of several different intervention strategies for helping participants maintain weight loss over a period of 2½ years. This paper describes the process by which the WLM research group designed and implemented one of these maintenance programs, featuring an Internet website and an associated prompting system using automated email and telephone messages, and the lessons learned during that process. Effectiveness data will be reported in a separate paper.

Methods

Design of the Weight Loss Maintenance Trial

The WLM is a four-center, randomized clinical trial testing the long-term efficacy of different strategies for maintaining weight loss. The design of the WLM is described elsewhere [31]. Briefly, participants in the WLM started a 6-month initial weight loss program focused on reducing caloric intake and increasing moderate intensity physical activity. Those who lost 4 kg or more were then randomly assigned to either a no-further treatment control condition or to one of two active weight loss maintenance interventions. The maintenance interventions were a personal contact condition, in which participants were contacted monthly by a health counselor, and an interactive technology (IT) condition, in which participants were encouraged to use an interactive website designed to help them maintain their weight loss. Weight loss results from WLM will not be available until mid-2008. The purpose of this paper is to describe the process by which the IT intervention program was developed, present utilization data for the first year, and provide a summary of what we learned from the development process.

Participants

We recruited adults with a body mass index (BMI) of 25-45 who were taking medication for either hypertension or hyperlipidemia. To be eligible, screening volunteers needed to have regular Internet and email access. Interested screening volunteers were sent an email message containing an individual identification number and the URL for a special screening website. Individuals needed to access that website and enter their identification number to be eligible for the study.

Weight Loss Interventions

This paper focuses on the development, implementation and use of one of the weight loss maintenance programs used in the WLM—the IT arm. Participants assigned to this arm used the website to record their weight, physical activity, and other weight loss activities. The website was designed to provide a number of important intervention elements, including social support using a bulletin board feature, record-keeping tools, tracking options, accountability, diet and exercise information, and tailored feedback.

Website Design

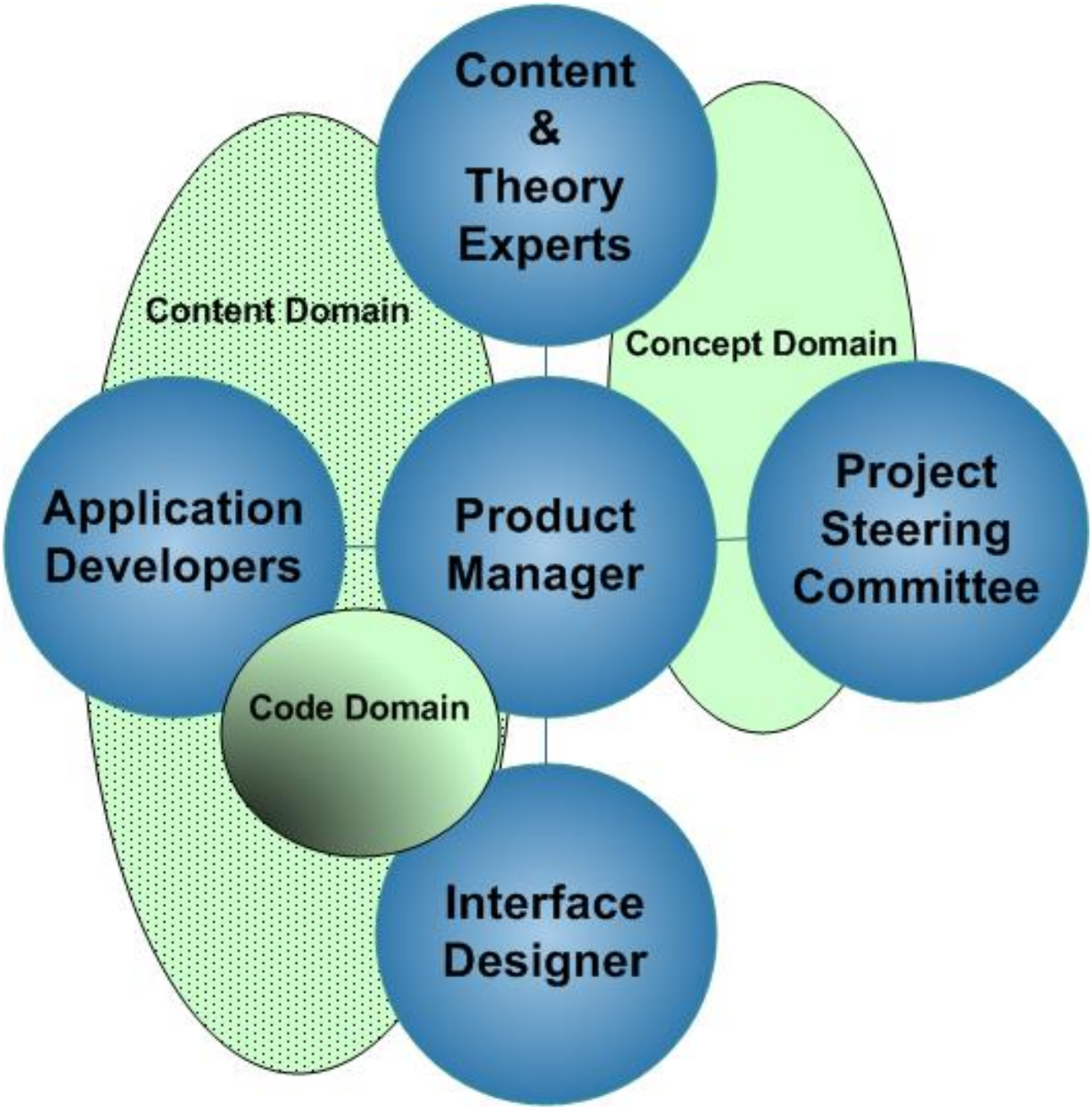
Successful design strategies for the study's interactive website can be summarized in three phases: (1) identifying the required skill sets for the design team, (2) specifying a stepwise process of designing the program, and (3) implementing the plan. Each phase is described in the following sections.

Design Team

Figure 1 displays an overview of the skills necessary for successfully designing an interactive behavior change website. Conceptual oversight, in this case the research project steering committee, determined the intervention's overall goals and theoretical framework. The committee specified the objectives and scope of the website, set priorities and timelines, and kept current with related activities in other aspects of the research project. The role of the steering committee was to "think big"

and provide scientific and conceptual guidance, but not to manage the project's day-to-day implementation. Oversight decisions set the course for website design and, once determined, can be costly to reverse. The plans made by this group are "big picture" decisions, and documentation of these decisions is critical to the forward progress of website design. Allowing form and function details to distract this group can be a major pitfall.

Figure 1. Overview of website design team; spheres and domains



WLM steering committee members included the principal investigators from a variety of specialty areas including psychology, cardiovascular health, epidemiology, nutrition, and clinical medicine. The theoretical framework chosen for this intervention combined self-directed behavior change theory [32,33], social support theory [34], motivational interviewing [35], and the transtheoretical stages of change model [36].

In addition to determining the intervention's theoretical framework, the steering committee specified the overall objectives and scope of the website. They balanced the targeted outcome (in this case, using the website to maintain behaviors that promote maintenance of weight loss) with the available resources and the timeline for product development. Participants randomized to the maintenance phase of WLM had already achieved a minimum weight loss of 4 kg during a 20-session series of weekly group meetings. Thus, the website was designed for knowledgeable and successful participants with some experience in the application of behavior change techniques. Rather than building a website to prompt initial weight loss, our objective was to build a site to maintain and support existing behavior change habits while helping participants develop new self-management skills. Continued website activity by study participants has been identified as a concern in several website

intervention studies [28,37-39], and keeping participants engaged over a long follow-up period (2½ years) was a key consideration in our design process. Finally, the aspect of social support was highlighted in the website's overall objectives. Social support has been identified as a major supportive tool for continued behavior change [32,34]. Table 1 specifies the objectives of the WLM website.

Table 1. Objectives of the WLM interactive behavior change website

1. Reinforce existing behavioral self-management strategies
2. Facilitate and encourage new self-management skills
3. Improve self-efficacy for long-term weight management
4. Remain fresh and inviting to encourage regular, long-term contact
5. Promote social support among website users

Content and theory experts provide the scientific expertise necessary to translate the overall intervention goals into a website's interactive modules. Content experts for the WLM website included master interventionists with graduate-level training in psychology, health counseling, nutrition, and physical activity. Building on the conceptual work of the research scientists on the steering committee, content and theory experts took the design process to a more detailed level. The input from content and theory experts ensured that the overall objectives were met in ways that were consistent with behavior change principles. The WLM content experts used their experience in conducting in-person weight loss interventions to identify key features of effective counseling sessions. Once identified, these key features were translated into interactive modules. Examples of key counseling features included offering choice, providing feedback, facilitating commitments to goals and plans, and minimizing the role of information while maximizing the importance of self-management. This group also determined strategies for implementing website modules that focused on key weight management behaviors (ie, encouraging participants to weigh themselves at least weekly). Weight entry was a "gatekeeper" to the home page. If a participant did not enter a weight upon log-in, the system directed the participant to the weight entry screen, leaving all other features disabled until a weight was entered. In a face-to-face counseling session for weight maintenance, there is a clear expectation that a weight will be taken and discussed during the visit, and this pattern was used during the 20-week initial weight loss program in WLM. Thus, requiring entry of weight at least weekly was not expected to be a barrier to website use. Making additional data entry requirements, however, was seen as a potential barrier for frequent website use.

The content experts translated the intervention's objectives into specific plans and supplied most of the site content. The same background and skills needed to write an in-person curriculum are also needed when developing scripts for the interactive modules, but the automated systems have some constraints. For example, an in-person counseling session may be free flowing and touch on a variety of related and nonrelated issues before getting to the key counseling steps that move the participant toward a specific goal and action plan. Content for an interactive module, in contrast, must be at least conceived in a stepwise fashion. Because lifestyle change intervention counseling is an inherently iterative and tailored experience for both the counselor and the participant, developing and documenting content for use in an automated module is a challenge. See Table 2 for the steps used to develop interactive modules.

Table 2. Development of interactive modules

Step	Participant Task
1 Assess the situation	Identify the desired behavior change
2 Define the problem	Chose from a list of possible barriers
3 Determine a strategy	Decide on the best next step
4 Create a plan	Select one or more specific actions
5 Summarize and plan follow-up	Review a comprehensive plan and select a follow-up reminder date

The interface design specialists are the main contributors to the website's look and feel. They also establish user functionality guidelines to be applied consistently throughout the website. The consultative and programming knowledge contributed by a user interface design expert is highly valuable and cannot be overlooked as an essential component of effective website design. A well-conceived module, based in sound behavior change theory and written with intensely rich content, will be of no value if the user's experience is not considered during the design phase. Designing for a successful user experience considers details such as font style and size, balance of graphic and text, minimizing the "clicks" necessary to get to a desired place, and creating intuitive ways to navigate while simultaneously designing for wide variations in user hardware, software, and Internet service provider (ISP) limitations.

The application developers bring technical expertise unmatched by any other discipline. This role, simply stated, cannot be done by anyone but a skilled website developer. The developer group writes the website functionality code under the direction of the content experts. A labyrinth of behind-the-scenes systems support what appear to be simple modules. Multiple layers of

system checks, error reports, data logging, and security measures exist within the programming of an interactive module, much of which is never seen by the larger design team or by website users.

The product manager serves as the communications hub for the design team and coordinates the entire design effort. As an indispensable member of the design team, the product manager must possess a variety of diverse skills and interests. Our experience developing the WLM website provided invaluable insights into the skills required for successful product management. These skills include the ability to contribute to conceptual oversight conversations while also being technically proficient to manage and evaluate minute details. The product manager must be skilled enough in each design team group to translate ideas effectively between groups. Additionally, the product manager must have the authority to make and finalize decisions. In our case, it was very helpful that the product manager (Funk) was also a content expert.

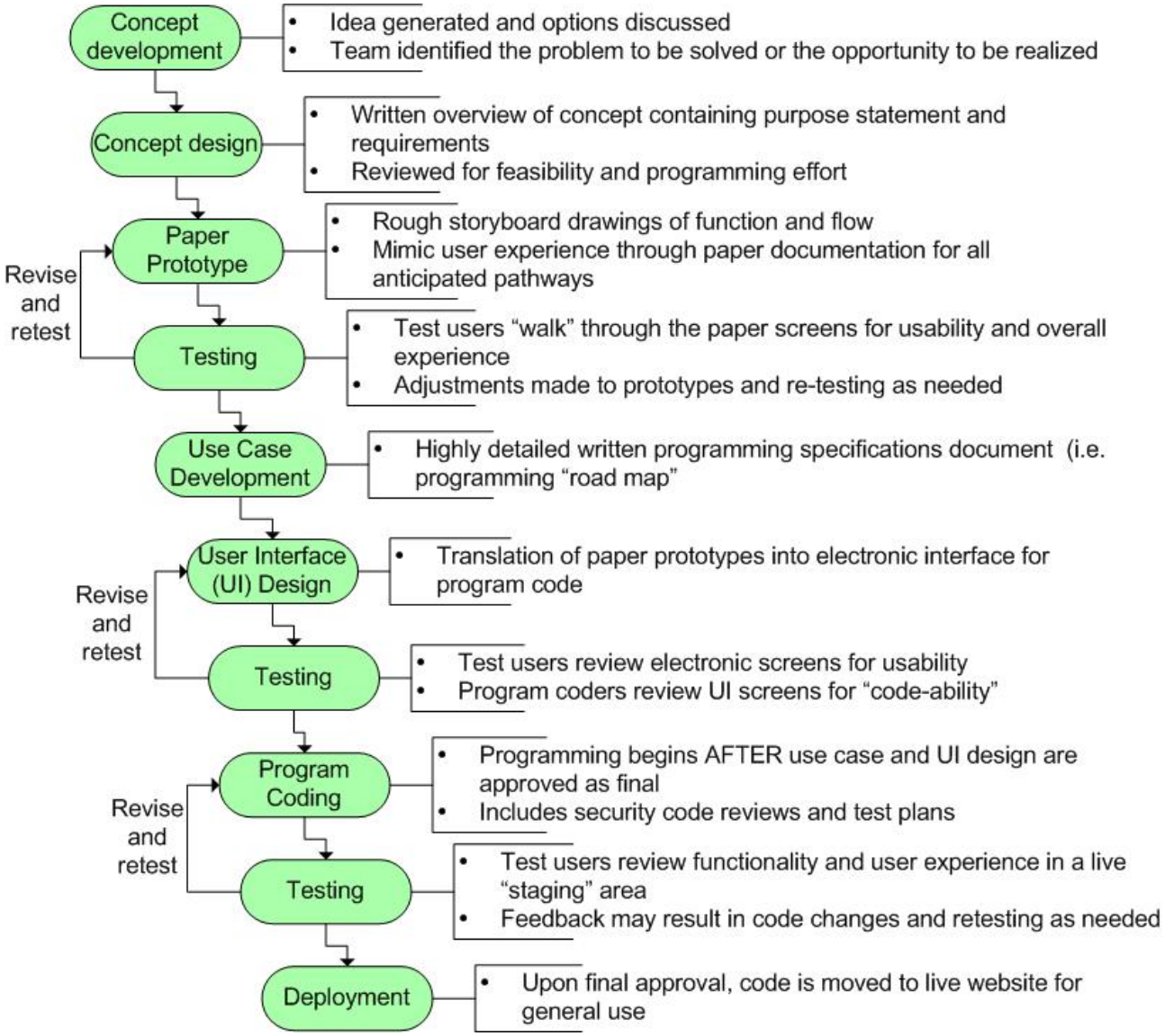
As with all teams, the groups of the website design team are interdependent. Whereas certain groups may interact frequently, others may interact only rarely. The "concept" domain (see Figure 1) includes the content and theory experts and the research scientists on the steering committee. A second domain, the "content" domain, includes the content and theory experts, applications developers, and user interface experts. A third "code" domain includes programmers and interface design experts. The product manager is a member of all domains. Other than the product manager's bridging work, the workings of the concept domain need not intersect with the roles and functions of the content and code domains. This is where the role of a product manager becomes essential. The product manager forms the bridge between the high-level overviews from the concept domain to the much more detailed, linear, and literal language used in the coding domain. Too much interaction between nonconnected domains can lead to confusion, rework, and possible team dissatisfaction. The product manager keeps the boundaries clear between domains and facilitates communication between the groups.

Design Process

The second phase of successful website design involves a clearly communicated stepwise process that outlines the development pathway of each interactive module. The design process described here assumes that the final website product is a unified collection of individually developed modules. The product manager maintains an ongoing record of each module's progress (outlined in Figure 2). There are several key factors involved in successfully moving through the website development process. The concept design step requires a written purpose that provides a "compass" for future design considerations.

The development of a paper prototype is an equally important design step. Despite the tediousness of drawing out the "what if" scenario for every possible pathway, the paper prototype process is essential for efficient work. We learned that easily modified paper prototypes highlight unresolved problems before expensive programming time has been invested. Finally, delaying program coding until final approval of the use case (the detailed programming specifications) is imperative. Our experience developing this website taught us that programmers code exactly according to specifications. The more specific the use case, the more likely the product will be what the originators envisioned. We instituted a "sign off" step whereby the product manager signed off on the use case document prior to any programming. We believe this step helped to minimize rework by holding the originators accountable to their specifications.

Figure 2. Steps in the website development process



Design Implementation

The WLM website's key interactive features include social support, self-monitoring, written guidelines for diet and physical activity, links to appropriate websites, supportive tools for behavior change, check-in accountability, tailored reinforcement messages, and problem solving and relapse prevention training. An overview of the key interactive features is shown in Figure 3. Our implementation plan included 3 months of beta testing. We received feedback from 44 pilot participants as well as project staff. Feedback included comments posted on the beta testing discussion forum using the website bulletin board, emailed comments to the website moderator, and comments solicited by phone. Pilot participants were asked to log in to the website at least weekly and use all the website features. They were not required to meet the study eligibility criteria; however, many used the site to help with their own personal weight control. Our main feedback objective was to understand the user experience and what would enhance utilization of the website features.

A typical log-in experience included a tailored welcome message and the option to enter weight and diet information before proceeding to the home page. At the home page, participants could choose to engage in any number of the interactive features listed in Figure 3 or to log out. Given the website's interactive nature and the intervention goal of weekly use for a 2½-year follow-up period, we learned from the beta testers that an individualized orientation to the website was much more likely to ensure user confidence and repeat log-ins than simply providing a website address and written instructions for use. Therefore, we instituted a participant orientation visit as part of our intervention protocol. Each participant was trained to use the website during an individual visit with a WLM staff interventionist. This training included an account setup during which the participant chose a display name, a first time check-in that demonstrated the usual weekly check-in expectations, and time to practise navigating the different features of the website, including an opportunity to post a message on the bulletin board. While the WLM participants had to pass a simple screening test for Internet access (receiving an email and visiting a special screening website), interventionists were trained to watch for specific technical barriers and to counsel accordingly during

the orientation. Sample screenshots of the website home page and participant's goal setting page are shown in Figure 4 and Figure 5, respectively. For a full overview of the WLM website screens, see the Multimedia Appendix.

Figure 3. Overview of the WLM key interactive features

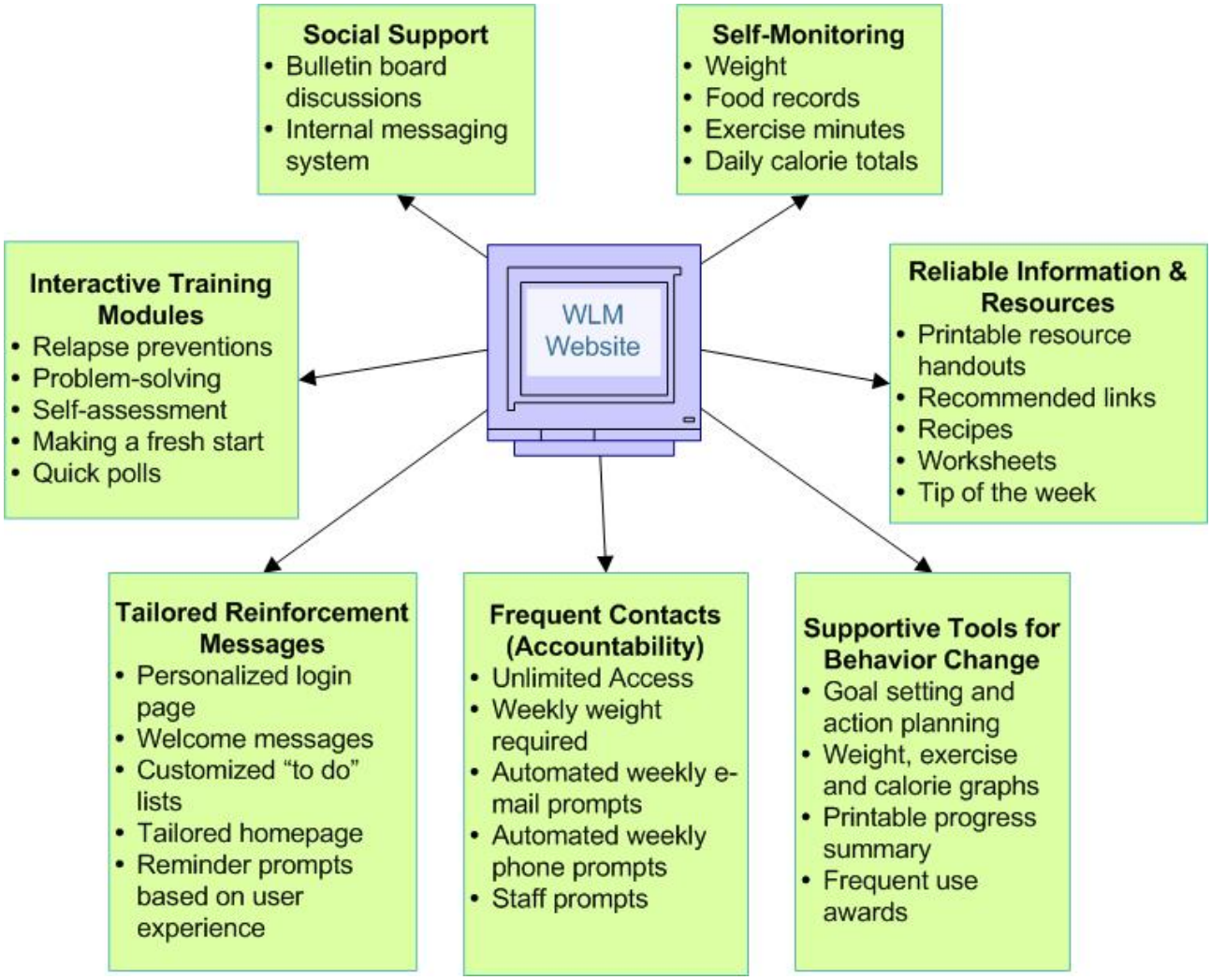


Figure 4. Sample screenshot of the website home page

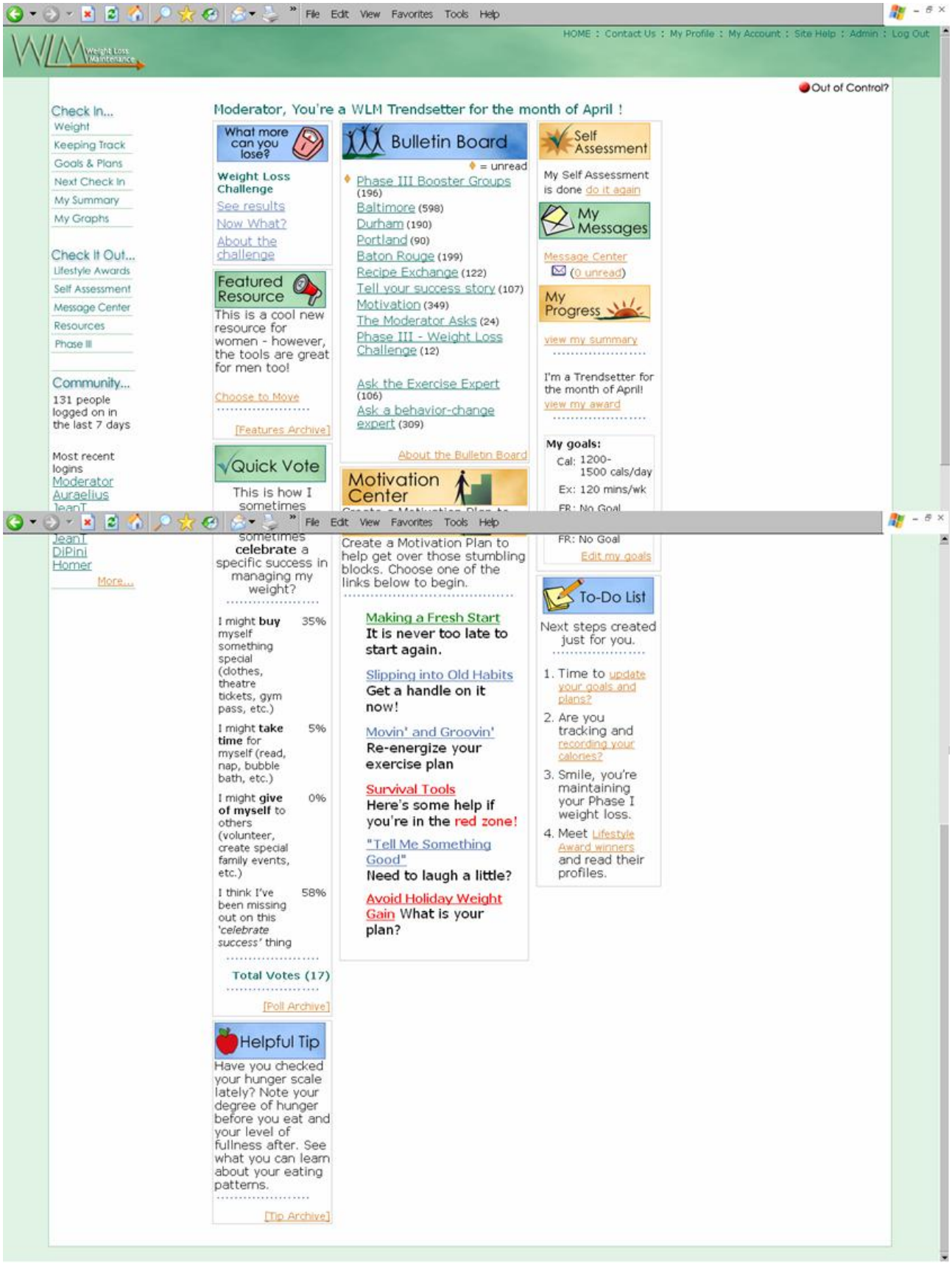
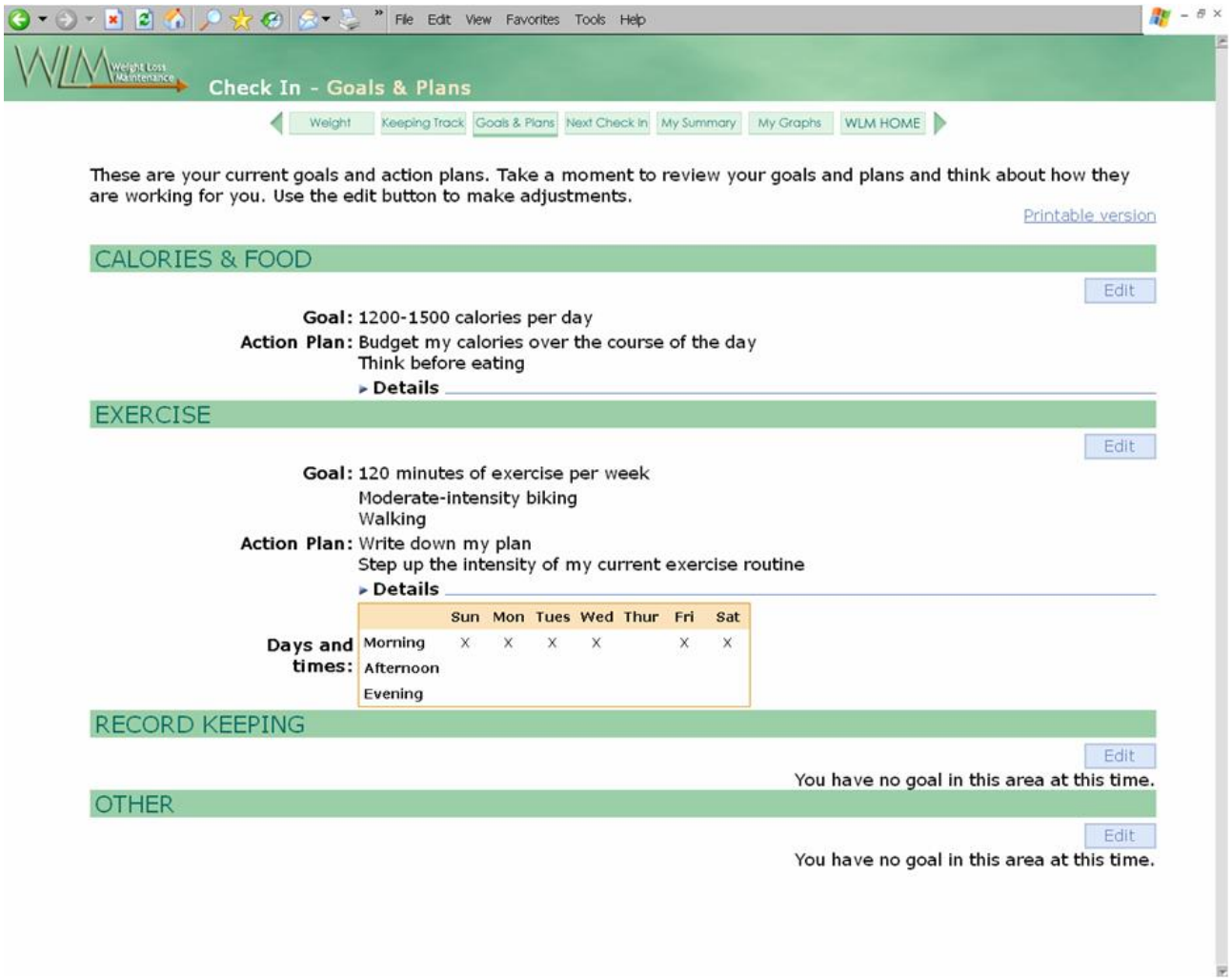


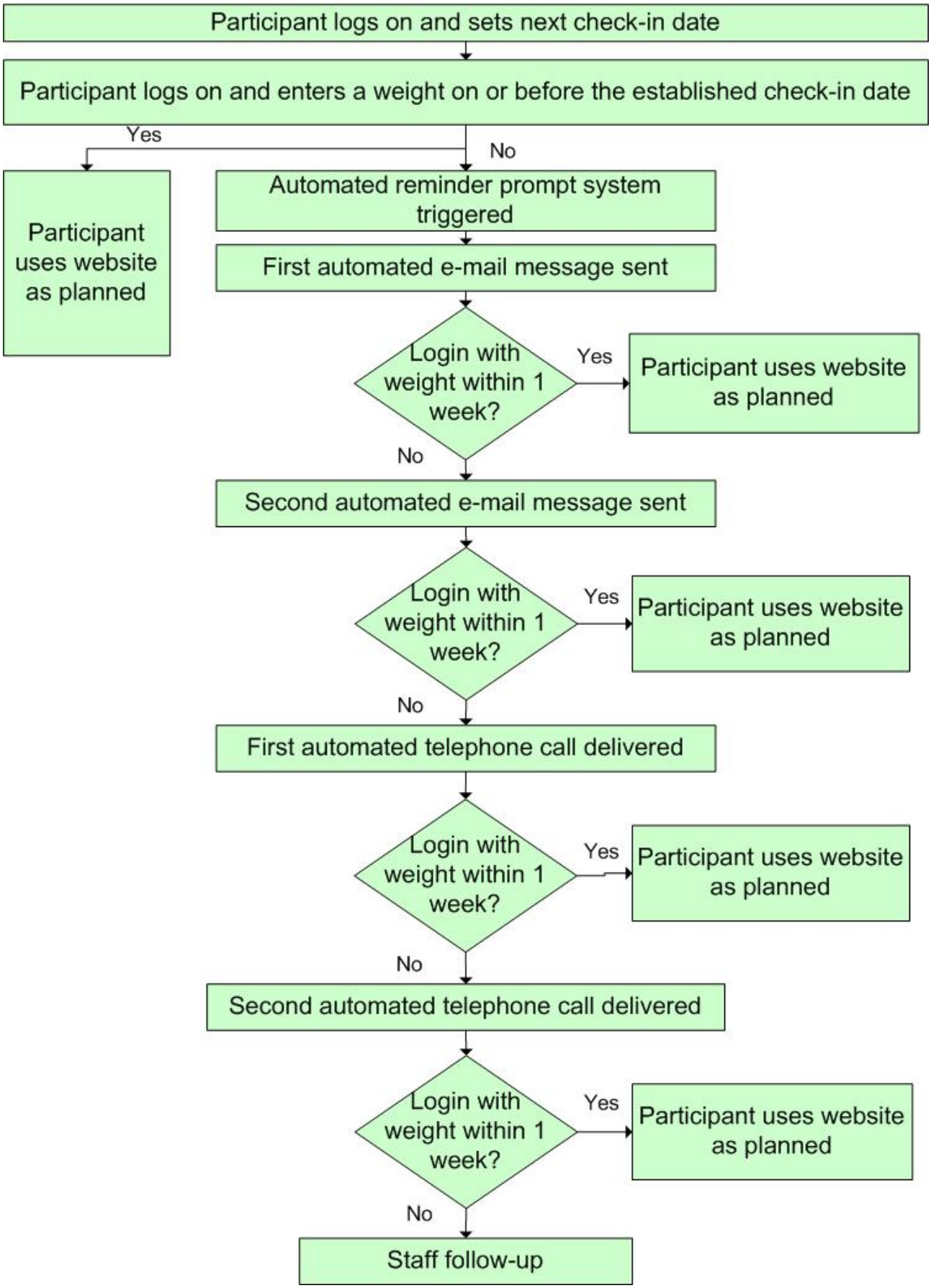
Figure 5. Sample screenshot of a participant's goal setting page



Prompting Use

Our weight loss maintenance program included an automated reminder system that prompted participants to return to the website if they missed their check-in date. An overview of the reminder system is shown in Figure 6. Specifically, if a participant had not logged on to the website on or before their next weekly check-in date, an automated email reminder was sent. Those participants who did not log on to the website within 1 week were sent a second email message prompt. Both email prompts were personalized, written in the spirit of motivational enhancement counseling (offering choice rather than instruction or advice), and contained a direct link to the website. If participants did not log on to the website within 1 week of the second email prompt, we employed automated telephone technology. This phone message was personalized and encouraged the participant to return to the website. A second automated telephone message was sent if the participant did not log in within 1 week of the first automated call. The system made every attempt to deliver the automated phone message, including multiple tries if the line was busy or hang-ups occurred. If there was no log-in response to the email and telephone automated prompts, a staff member called the participant. This individual effort continued until either the participant logged on to the website or the study ended.

Figure 6. Overview of automated prompt system



Results

Table 3 shows the characteristics of the 348 WLM participants assigned to the IT arm of the trial; 63% were female and 38% were African American. The mean age was 56 years, and the mean BMI at the start of the initial weight loss intervention was 34. IT participants lost a mean of 9 kg during the initial weight loss treatment.

Table 3. WLM IT participant characteristics (N = 348)

Mean age (years)	56
% female	63
% African American	38
Mean BMI at start of phase I	34
Mean weight loss in phase I (kg)	9

Weight loss data from the trial will not be available until mid-2008, but preliminary data on website use during the first year following randomization are presented here. During the first year, active website use (defined as at least one log-in per month) remained high, with over 80% of the participants still using the website in month 12 (Figure 7). In other words, approximately 20% of participants were no longer active users of the website after 12 months of the intervention. Furthermore, during the first 52 weeks, participants averaged 35 weeks with at least one log-in.

Figure 7. Percent of participants with at least one log-in per month

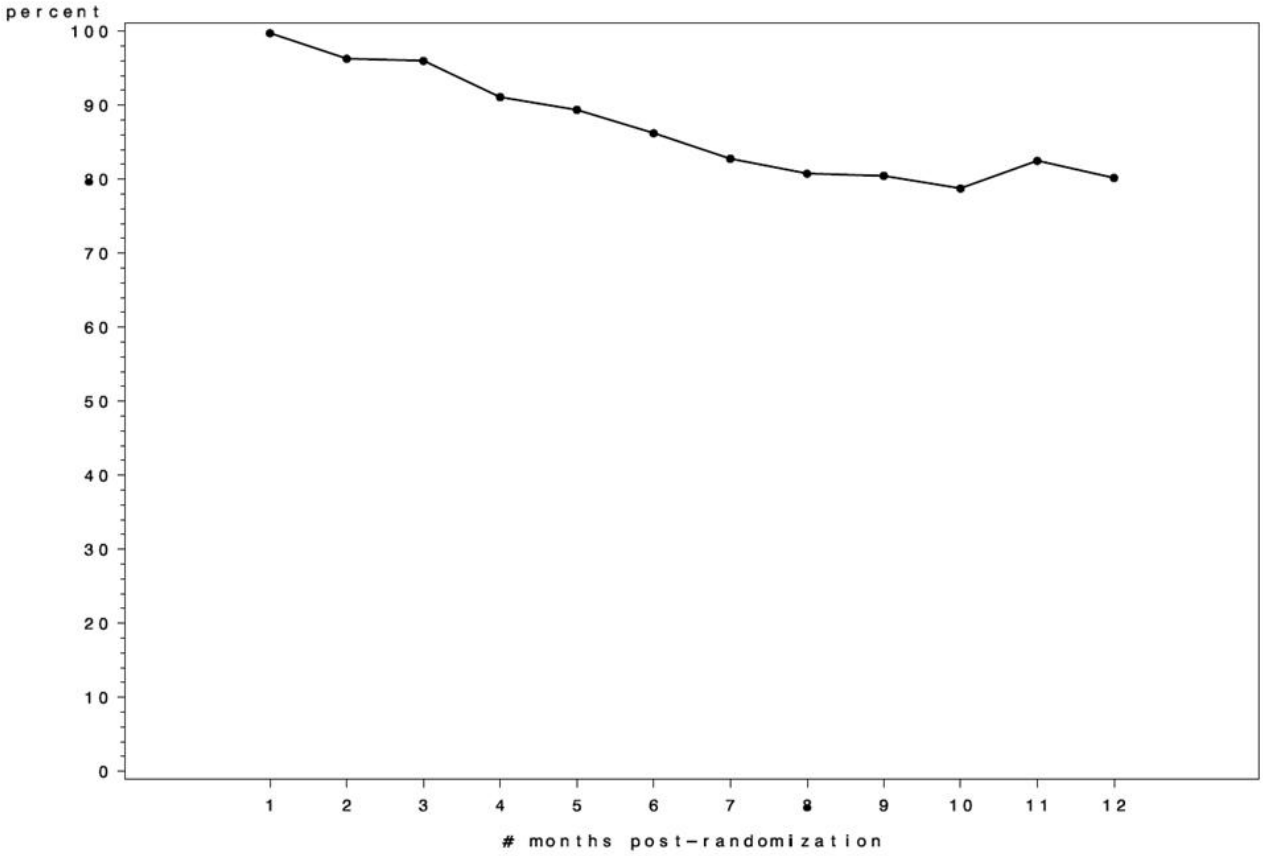


Table 4 shows the cumulative percentage of log-ins occurring as a result of the automated reminder prompt system: 86.4% of the first weekly email prompts resulted in the participant logging in to the website within 1 week, and 56.7% of the second email prompts resulted in a log-in within 1 week. Overall, the escalating series of email and automated telephone calls effectively prompted participants to return to the website in 97.3% of cases.

Table 4. Log-ins as a result of the automated prompts (first 12 months)

Prompt	No.	Logged In After Prompt (%)	Cumulative Percent
1st auto email	9372	86.4	86.4
2nd auto email	1278	56.7	94.1
1st auto phone call	554	36.8	96.3
2nd auto phone call	350	26.9	97.3
Staff phone call	256	64.8	99.0
Never responded	90	0	100.0

Discussion

Behavior change websites not only offer lower cost per participant than face-to-face contacts with counselors, but also hold promise as an unexplored new mechanism for supporting long-term behavior change. Our experience developing the WLM behavior change website shows that high rates of use can be maintained for at least 1 year. Participation in this study was much greater than in some other long-term behavioral interventions [40-42], but similar to that seen in the Stop-Regain trial [43].

We have identified several major lessons from our experience developing and implementing the WLM interactive website. First, it is essential to specify the theoretical foundation of the intervention program and the website objectives early in the design process. Website design does not iterate the same way as development of in-person, counseling-based interventions. Making clear decisions about intervention objectives and abiding by them during the design process helps eliminate costly rework.

Our second key learning was that detailed paper prototypes and specification documents should always precede programming. While the design team may want to jump directly to screenshots before the logic has been thoroughly outlined, the results tend to be better if they wait for the paper prototyping to be completed. In the WLM, we called paper prototype meetings "wall meetings" because hours were spent taping freehand paper "screens" to the wall and determining the outcome of every link. These wall meetings resulted in many modifications and intervention improvements that would have been difficult during later stages in the development process. The final set of freehand paper screens was also very useful when writing the detailed specifications (use case) for the programmer.

A final key learning was to not underestimate the essential role of a product manager. To have all groups doing what they do best requires a central team member to intersect with each group. We believe that the product manager must be able to manage in all three domains (concept, content, and code) to ensure an effective website design process. The product manager helps to keep the development team working toward a common goal while serving as a translator for those working at different levels of detail.

Additionally, we gained insight into participant use of an interactive website and what is required to keep participants engaged. Based on preliminary "hit" counts (data not shown), interactive features like the weight entry form and the bulletin board discussion appeared to be most popular. We were concerned that a website intervention would have problems maintaining the interest of participants for long-term follow-up. We found, however, that a high rate of participation can be achieved for at least 12 months with an interactive website. The automated email and telephone reminders were quite effective in prompting regular use of the website for at least 1 year. In an 8-week, stage-based physical activity website intervention, Leslie et al [39] determined that emails prompted return visits to a website (77% returned after email prompts), but that the same emails were not helpful for encouraging new users to visit for the first time. In contrast to the Leslie et al study, participants in WLM were initially oriented to the website through a 1-hour individual visit with an interventionist. Personalized website orientation may be a critical factor in the effectiveness of subsequent email prompts that encourage returning to the website. Another study [44] reported that well-constructed email messages can have a beneficial effect on diet and exercise health behaviors. The authors suggest, however, that the email messages may need further tailoring and grounding in health behavior change theory to strengthen their potential. The WLM email prompts were customized to the individual user, easy to read, provided choice rather than advice, and included a link for easy access to the WLM website. These factors were potentially positive contributors to the effectiveness of our email prompts. We were initially concerned about the number and frequency of email reminders sent to study participants. We only sent email prompts when the participant met the specified criteria of not logging on to the website within 1 week of the last log-in. However, during the first 6 months of the intervention, 83% of participants received a weekly email prompt to log on to the website, providing us with two important lessons: (1) participants are not bothered by reminders to return to the website, and (2) in general, participants do not set up outside reminders to log in, but simply wait to be prompted.

Finally, it is important to note the current limitations of Web-based programs. Developers of behavior change websites must be prepared to continually update the product and limit the use of available technologies in consideration of bandwidth limitations. Danaher et al [45] urge developers to consider the bandwidth necessary to operate rich-media websites as a possible barrier to participant use. User frustration resulting from long page downloads presents a near terminal problem for researchers looking to test behavioral website use. Therefore, we limited the bandwidth requirements of the WLM website to accommodate those with limited bandwidth. For this reason, the WLM website is devoid of photos, moving text, video clips,

and music. Also acknowledged by Danaher et al [45], the scalability of a behavior change website must be considered at the time of development. The capacity of a website to "grow" beyond its current capacity is an essential consideration.

Given our study timeline, we developed and implemented this website in 12 months. Looking back on our experience, additional development time would have been beneficial in three key processes: (1) at least 6 months of general beta testing (we had 3 months), (2) an even richer understanding of the user experience from the pilot participants (ie, periodic individual interviews to understand how/if the user experience evolved over time), and (3) additional opportunities to test multiple prompting strategies for encouraging participants to continue using the site. Even without such additional development time, use of this website remained high throughout the first year, with over 80% of the participants continuing to be active users during the 12th month.

Acknowledgments

This study was supported by National Heart, Lung, and Blood Institute grants 5-U01 HL68920, 5-U01 HL68734, 5-U01 HL68676, 5-U01 HL68790, and 5- HL68955.

Conflicts of Interest

None declared.

Multimedia Appendix

Sample screenshots of the WLM interactive website

[PPT file (MS Powerpoint), 3 MB - jmir_v10i1e1_app1.ppt]

References

1. Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. JAMA 2003 Jan 1;289(1):76-79 [FREE Full text] [Medline: 12503980] [doi: 10.1001/jama.289.1.76]
2. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. JAMA 2006 Apr 5;295(13):1549-1555 [FREE Full text] [Medline: 16595758] [doi: 10.1001/jama.295.13.1549]
3. Wang Y, Beydoun MA. The obesity epidemic in the United States--gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. Epidemiol Rev 2007;29(1):6-28. [Medline: 17510091] [doi: 10.1093/epirev/mxm007]
4. Blanck HM, Dietz WH, Galuska DA, et al. State-specific prevalence of obesity among adults - United States 2005. MMWR Morb Mortal Wkly Rep 2006;55:985-988.
5. Seidell JC, Visscher TL, Hoogeveen RT. Overweight and obesity in the mortality rate data: current evidence and research issues. Med Sci Sports Exerc 1999 Nov;31(11 Suppl):S597-S601. [Medline: 10593534]
6. Allison DB, Fontaine KR, Manson JE, Stevens J, VanItallie TB. Annual deaths attributable to obesity in the United States. JAMA 1999 Oct 27;282(16):1530-1538 [FREE Full text] [Medline: 10546692] [doi: 10.1001/jama.282.16.1530]
7. Fontaine KR, Redden DT, Wang C, Westfall AO, Allison DB. Years of life lost due to obesity. JAMA 2003 Jan 8;289(2):187-193 [FREE Full text] [Medline: 12517229] [doi: 10.1001/jama.289.2.187]
8. Olshansky SJ, Passaro DJ, Hershow RC, Layden J, Carnes BA, Brody J, et al. A potential decline in life expectancy in the United States in the 21st century. N Engl J Med 2005 Mar 17;352(11):1138-1145. [Medline: 15784668] [doi: 10.1056/NEJMsr043743]
9. Haslam DW, James WPT. Obesity. Lancet 2005 Oct 1;366(9492):1197-1209. [Medline: 16198769] [doi: 10.1016/S0140-6736(05)67483-1]
10. Allison DB, Zannolli R, Narayan KM. The direct health care costs of obesity in the United States. Am J Public Health 1999 Aug;89(8):1194-1199 [FREE Full text] [Medline: 10432905]
11. Thompson D, Brown JB, Nichols GA, Elmer PJ, Oster G. Body mass index and future healthcare costs: a retrospective cohort study. Obes Res 2001 Mar;9(3):210-218 [FREE Full text] [Medline: 11323447]
12. Musich S, Lu C, McDonald T, Campagne LJ, Edington DW. Association of additional health risks on medical charges and prevalence of diabetes within body mass index categories. Am J Health Promot 2004;18(3):264-268. [Medline: 14748319]
13. Finkelstein EA, Fiebelkorn IC, Wang G. State-level estimates of annual medical expenditures attributable to obesity. Obes Res 2004 Jan;12(1):18-24 [FREE Full text] [Medline: 14742838]
14. ; National Heart, Lung and Blood Institute. Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report. Executive Summary. Washington, DC: US Department of Health and Human Services; 1998.
15. Wadden TA, Crerand CE, Brock J. Behavioral treatment of obesity. Psychiatr Clin North Am 2005 Mar;28(1):151-70, ix. [Medline: 15733617] [doi: 10.1016/j.psc.2004.09.008]

16. Wing RR, Hill JO. Successful weight loss maintenance. Annu Rev Nutr 2001;21:323-341. [Medline: 11375440] [doi: 10.1146/annurev.nutr.21.1.323]
17. Wadden TA, Butryn ML, Byrne KJ. Efficacy of lifestyle modification for long-term weight control. Obes Res 2004 Dec;12 Suppl:151S-162S [FREE Full text] [Medline: 15687411]
18. ; The Trials of Hypertension Prevention Collaborative Research Group. Effects of weight loss and sodium reduction intervention on blood pressure and hypertension incidence in overweight people with high-normal blood pressure. The Trials of Hypertension Prevention, Phase II. Arch Intern Med 1997;157(6):657-667. [doi: 10.1001/archinte.157.6.657]
19. Stevens VJ, Obarzanek E, Cook NR, Lee IM, Appel LJ, Smith West D, et al; Trials for the Hypertension Prevention Research Group. Long-term weight loss and changes in blood pressure: results of the Trials of Hypertension Prevention, phase II. Ann Intern Med 2001 Jan 2;134(1):1-11 [FREE Full text] [Medline: 11187414]
20. Appel LJ, Miller ER, Jee SH, Stolzenberg-Solomon R, Lin PH, Erlinger T, et al. Effect of dietary patterns on serum homocysteine: results of a randomized, controlled feeding study. Circulation 2000 Aug 22;102(8):852-857 [FREE Full text] [Medline: 10952952]
21. Wing RR, Hamman RF, Bray GA, Delahanty L, Edelstein SL, Hill JO, et al; Diabetes Prevention Program Research Group. Achieving weight and activity goals among diabetes prevention program lifestyle participants. Obes Res 2004 Sep;12(9):1426-1434 [FREE Full text] [Medline: 15483207]
22. Perri MG, Nezu AM, Patti ET, McCann KL. Effect of length of treatment on weight loss. J Consult Clin Psychol 1989 Jun;57(3):450-452. [Medline: 2500466] [doi: 10.1037/0022-006X.57.3.450]
23. Wing RR, Phelan S. Long-term weight loss maintenance. Am J Clin Nutr 2005 Jul;82(1 Suppl):222S-225S [FREE Full text] [Medline: 16002825]
24. Jeffery RW, Drewnowski A, Epstein LH, Stunkard AJ, Wilson GT, Wing RR, et al. Long-term maintenance of weight loss: current status. Health Psychol 2000 Jan;19(1 Suppl):5-16. [Medline: 10709944] [doi: 10.1037/0278-6133.19.Suppl1.5]
25. Eng TR, Gustafson DH, E. Science Panel on Interactive Communication and Health. Wired for Health and Well-Being: The Emergence of Interactive Health Communication. Washington, DC: US Government Printing Office; 1999:1-41.
26. Revere D, Dunbar PJ. Review of computer-generated outpatient health behavior interventions: clinical encounters "in absentia". J Am Med Inform Assoc 2001;8(1):62-79 [FREE Full text] [Medline: 11141513]
27. Jimison H, Adler L, Coye M, Mulley A, Eng TR. Health care providers and purchasers and evaluation of interactive health communication applications. Science Panel on Interactive Communication and Health. Am J Prev Med 1999 Jan;16(1):16-22. [Medline: 9894550] [doi: 10.1016/S0749-3797(98)00105-6]
28. Harvey-Berino J, Pintauro S, Buzzell P, Gold EC. Effect of internet support on the long-term maintenance of weight loss. Obes Res 2004 Feb;12(2):320-329 [FREE Full text] [Medline: 14981225]
29. Harvey-Berino J, Pintauro SJ, Gold EC. The feasibility of using Internet support for the maintenance of weight loss. Behav Modif 2002 Jan;26(1):103-116. [Medline: 11799651] [doi: 10.1177/0145445502026001006]
30. Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11 [FREE Full text] [Medline: 17513282] [doi: 10.2196/jmir.9.2.e11]
31. Brantley PJ, Appel LJ, Hollis JF, et al. Weight Loss Maintenance (WLM): design and rationale of a multi-center trial to sustain weight loss. Contemp Clin Trials (forthcoming).
32. Watson DL, Tharp RG. Self-Directed Behavior: Self-Modification for Personal Adjustment. 8th edition. Belmont, CA: Wadsworth Publishing; 2002.
33. Martin GL, Pear J. Behavior Modification: What Is It and How to Do It. 7th edition. New York: Prentice Hall; 2003.
34. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [Medline: 15090118] [doi: 10.1177/1090198104263660]
35. Miller WR, Rollnick S. Motivational Interviewing: Preparing People for Change. 2nd edition. New York: The Guilford Press; 2002.
36. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [Medline: 10170434]
37. Tate DF, Jackvony EH, Wing RR. Effects of Internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. JAMA 2003 Apr 9;289(14):1833-1836 [FREE Full text] [Medline: 12684363] [doi: 10.1001/jama.289.14.1833]
38. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 7;285(9):1172-1177 [FREE Full text] [Medline: 11231746] [doi: 10.1001/jama.285.9.1172]
39. Leslie E, Marshall AL, Owen N, Bauman A. Engagement and retention of participants in a physical activity website. Prev Med 2005 Jan;40(1):54-59. [Medline: 15530581] [doi: 10.1016/j.ypmed.2004.05.002]
40. Rothert K, Strecher VJ, Doyle LA, Caplan WM, Joyce JS, Jimison HB, et al. Web-based weight management programs in an integrated health care setting: a randomized, controlled trial. Obesity (Silver Spring) 2006 Feb;14(2):266-272 [FREE Full text] [Medline: 16571852]

41. Formica M, Kabbara K, Clark R, McAlindon T. Can clinical trials requiring frequent participant contact be conducted over the Internet? Results from an online randomized controlled trial evaluating a topical ointment for herpes labialis. J Med Internet Res 2004 Feb 17;6(1):e6 [FREE Full text] [Medline: 15111272] [doi: 10.2196/jmir.6.1.e6]
42. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a Web-based self-help cognitive behavioral therapy program for panic disorder. J Med Internet Res 2005;7(1):e7 [FREE Full text] [Medline: 15829479] [doi: 10.2196/jmir.7.1.e7]
43. Wing RR, Tate DF, Gorin AA, Raynor HA, Fava JL. A self-regulation program for maintenance of weight loss. N Engl J Med 2006 Oct 12;355(15):1563-1571 [FREE Full text] [Medline: 17035649] [doi: 10.1056/NEJMoa061883]
44. Plotnikoff RC, McCargar LJ, Wilson PM, Loucaides CA. Efficacy of an E-mail intervention for the promotion of physical activity and nutrition behavior in the workplace context. Am J Health Promot 2005;19(6):422-429. [Medline: 16022206]
45. Danaher BG, Jazdzewski SA, McKay HG, Hudson CR. Bandwidth constraints to using video and other rich media in behavior change websites. J Med Internet Res 2005;7(4):e49 [FREE Full text] [Medline: 16236701] [doi: 10.2196/jmir.7.4.e49]

Abbreviations

BMI: body mass index
IT: interactive technology condition
WLM: Weight Loss Maintenance Trial

submitted 19.04.07; peer-reviewed by N Sherwood, J Wright, R Bensley; comments to author 24.07.07; revised version received 28.11.07; accepted 04.01.08; published 25.01.08.

Please cite as:
Stevens VJ, Funk KL, Brantley PJ, Erlinger TP, Myers VH, Champagne CM, Bauck A, Samuel-Hodge CD, Hollis JF
Design and Implementation of an Interactive Website to Support Long-Term Maintenance of Weight Loss
J Med Internet Res 2008;10(1):e1
URL: http://www.jmir.org/2008/1/e1/
doi:10.2196/jmir.931
PMID:18244892

Original Paper

Effectiveness of a Web-Based Self-Help Intervention for Symptoms of Depression, Anxiety, and Stress: Randomized Controlled Trial

Annemieke van Straten[1], PhD; Pim Cuijpers[1], PhD; Niels Smits[1], PhD

Vrije Universiteit Amsterdam, Department of Clinical Psychology, Amsterdam, The Netherlands

Corresponding Author:
Annemieke van Straten, PhD
FPP, Department of Clinical Psychology
Vrije Universiteit Amsterdam
Van der Boechorststraat 1
1081 BT Amsterdam
The Netherlands
Phone: +31 20 5988970
Fax: +31 20 5988758
Email: a.van.straten@psy.vu.nl

Abstract

Background: Self-help therapies are often effective in reducing mental health problems. We developed a new Web-based self-help intervention based on problem-solving therapy, which may be used for people with different types of comorbid problems: depression, anxiety, and work-related stress.

Objective: The aim was to study whether a Web-based self-help intervention is effective in reducing depression, anxiety, and work-related stress (burnout).

Methods: A total of 213 participants were recruited through mass media and randomized to the intervention (n = 107) or a waiting list control group (n = 106). The Web-based course took 4 weeks. Every week an automated email was sent to the participants to explain the contents and exercises for the coming week. In addition, participants were supported by trained psychology students who offered feedback by email on the completed exercises. The core element of the intervention is a procedure in which the participants learn to approach solvable problems in a structured way. At pre-test and post-test, we measured the following primary outcomes: depression (CES-D and MDI), anxiety (SCL-A and HADS), and work-related stress (MBI). Quality of life (EQ-5D) was measured as a secondary outcome. Intention-to-treat analyses were performed.

Results: Of the 213 participants, 177 (83.1%) completed the baseline and follow-up questionnaires; missing data were statistically imputed. Of all 107 participants in the intervention group, 9% (n = 10) dropped out before the course started and 55% (n = 59) completed the whole course. Among all participants, the intervention was effective in reducing symptoms of depression (CES-D: Cohen's d = 0.50, 95% confidence interval (CI) 0.22-0.79; MDI: d = 0.33, 95% CI 0.03-0.63) and anxiety (SCL-A: d = 0.42, 95% CI 0.14-0.70; HADS: d = 0.33, 95% CI 0.04-0.61) as well as in enhancing quality of life (d = 0.31, 95% CI 0.03-0.60). Moreover, a higher percentage of patients in the intervention group experienced a significant improvement in symptoms (CES-D: odds ratio [OR] = 3.5, 95% CI 1.9-6.7; MDI: OR = 3.7, 95% CI 1.4-10.0; SCL-A: OR = 2.1, 95% CI 1.0-4.6; HADS: OR = 3.1, 95% CI 1.6-6.0). Patients in the intervention group also recovered more often (MDI: OR = 2.2; SCL-A: OR = 2.0; HADS < 8), although these results were not statistically significant. The course was less effective for work-related stress, but participants in the intervention group recovered more often from burnout than those in the control group (OR = 4.0, 95% CI 1.2-13.5).

Conclusions: We demonstrated statistically and clinically significant effects on symptoms of depression and anxiety. These effects were even more pronounced among participants with more severe baseline problems and for participants who fully completed the course. The effects on work-related stress and quality of life were less clear. To our knowledge, this is the first trial of a Web-based, problem-solving intervention for people with different types of (comorbid) emotional problems. The results are promising, especially for symptoms of depression and anxiety. Further research is needed to enhance the effectiveness for work-related stress.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 14881571

(J Med Internet Res 2008;10(1):e7) doi:10.2196/jmir.954

KEYWORDS

Bibliotherapy; psychotherapy; problem-solving therapy; depression; anxiety; stress

Introduction

It has been convincingly demonstrated that self-help therapies are effective in reducing mental health problems [1-5]. A self-help therapy can be defined as a standardized psychological treatment that the patient works through independently at home [6]. It is commonly delivered in book format, in which case it is called "bibliotherapy." However, the therapy can also be delivered through other media, such as CD-ROMs, television programs, or videotapes. In recent years, self-help has been increasingly offered through the Internet [5,7,8]. Web-based self-help may be an effective and inexpensive alternative to more traditional therapies, especially since the majority of persons in the general population with a mental health disorder (an estimated 65%) do not receive help from any professional mental health services [9,10].

The self-help therapies that are currently available have all been developed for patients with a specific disorder, such as depression, panic disorder, social phobia, general anxiety disorder, or posttraumatic stress disorder, and most are based on cognitive behavioral therapy. Problem-solving therapy, a brief form of psychotherapy where patients identify their most immediate problems and ways of regaining control over them, are not limited to one specific disorder and may be effective in several problem areas. Face-to-face problem-solving therapies have been shown to be effective in depression [11,12] and several other mental health problems [13-15]. We know that at least one Web-based cognitive behavioral therapy includes a problem-solving module (MoodGYM) [16,17], but as far as we know, there is no Web-based therapy that uses problem solving as the core element. Therefore, we decided to develop a new, problem-focused, generic self-help method for multiple mental health problems that could be applied through the Internet.

As a general framework for the intervention, we used the model developed by Bowman and colleagues, which is based on problem-solving therapy [18,19]. The general idea of this intervention, which is called self-examination therapy, is that participants learn to regain control over their problems and lives by (1) determining what really matters to them, (2) investing energy only in those problems that are related to what matters, (3) thinking less negatively about the problems that are unrelated and, (4) accepting those situations that cannot be changed. This method has been found to be effective in several studies in the United States [14,19,20]. We used the self-examination therapy as a framework for our intervention but translated it into Dutch, elaborated on it, and added information and exercises. We built a website for this intervention and developed a system for email support.

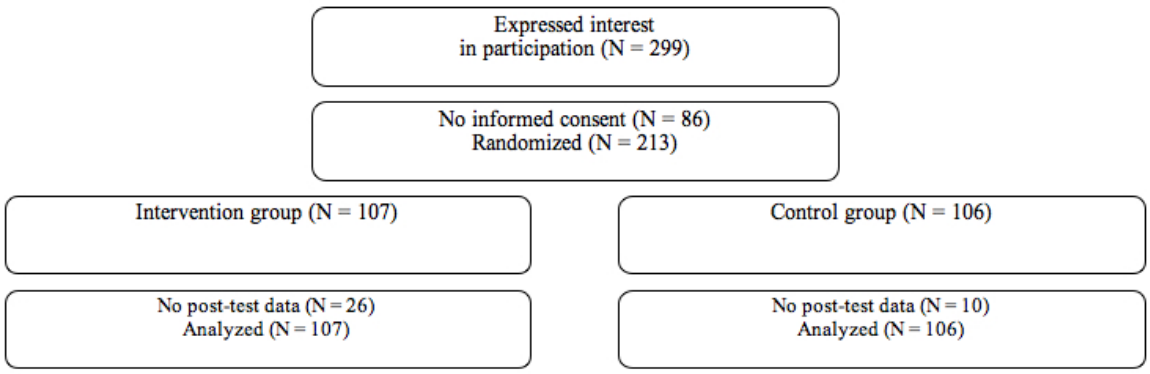
The aim of this study was to determine the effectiveness of this Web-based generic treatment method for participants with depression, anxiety, and work-related stress.

Methods

Recruitment of Participants

We recruited participants through advertisements about Internet self-help treatment for symptoms of depression, anxiety, and work-related stress placed in local and national newspapers. We aimed at including 200 participants in order to be able to demonstrate moderate effects of $d = 0.40$ while using a power $(1 - \beta)$ of 80% and an alpha of .05. We were contacted through email by 299 people (Figure 1). These 299 potential participants received an information booklet and an informed consent form by post as well as a baseline questionnaire through the Internet. All 213 individuals who returned the informed consent and the baseline questionnaire were included. No inclusion or exclusion criteria were used because the intervention was aimed at the general population. Enrollment took place between November 30 and December 20, 2005. The study was approved by the Medical Ethical Committee of the Vrije Universiteit Medical Center, Amsterdam, The Netherlands.

Figure 1. Flowchart of participants



Intervention

The intervention was Web-based (see Multimedia Appendix for screenshots). Participants were provided with a username and password to access the website. Every week an automated email was sent to the participants to explain the contents and exercises for the coming week. All the information as well as the exercise forms could also be downloaded from the website in case participants preferred to read the information on paper. Master's level psychology students, trained and supervised by the authors (PC, AvS), offered feedback on the completed exercises. This feedback was not therapeutic but was directed at mastering the proposed problem-solving strategies. For a participant completing the course, the total time spent by the psychology students on feedback was approximately 45 minutes. The course takes 4 weeks.

The intervention consists of three steps:

1. Participants describe what really matters to them.
2. Participants write down their current worries and problems and categorize them into three types: (a) unimportant problems (problems unrelated to the things that matter to them), (b) problems that can be solved, and (c) problems that cannot be solved (eg, the loss of a loved one).
3. Participants make a plan for the future in which they describe how they will try to accomplish those things that matter most to them.

The second step is the most important of the intervention. For each of the three types of problem (ie, a, b, and c), a different strategy is proposed to cope with it. For the solvable problems (ie, b), we propose the following procedure: (1) write a clear definition of the problem, (2) generate multiple solutions to the problem, (3) select the best solution, (4) work out a systematic plan for this solution, (5) carry out the solution, and (6) evaluate as to whether the solution has resolved the problem.

Design

All participants were randomly assigned to either the self-help course or a waiting list. Questionnaires were sent before the start of the course and 5 weeks later, after the intervention group had finished. Thereafter, the participants in the waiting list group could complete the course.

Randomization

Randomization took place 1 week before the start of the intervention. We used block randomization with blocks of 10. The randomization scheme was derived by computer and carried out by an independent researcher. All participants were informed by email about the randomization outcome.

Measures

Depressive symptoms were measured with the Center for Epidemiological Studies Depression Scale (CES-D) [21] and the Major Depression Inventory (MDI) [22]. The CES-D is a 20-item, self-report questionnaire on feelings of depression; its total score ranges from 0 (no depressive symptoms at all) to 60 (many depressive symptoms). The MDI contains 12 items that are used to calculate the scores on the 10 ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th Revision) symptoms of depression. Each of the 10 symptoms is scored on a scale from 0 (at no time) to 5 (all of the time). The total score is calculated by adding all the items, and thus ranges from 0 to 50. Based on the symptom scores, it is also possible to determine the presence or absence of major depression according to the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th Edition) criteria.

Symptoms of anxiety were measured with the seven anxiety questions of the Hospital Anxiety and Depression Scale (HADS) [23] and the anxiety section of the Symptom Checklist (SCL-A) [24]. The total score of the HADS varies from 0 (no complaints of anxiety) to 21 (many complaints of anxiety). The SCL-A consists of 10 questions, and the total score ranges from 10 (no complaints of anxiety) to 50 (many complaints of anxiety).

Work-related stress was measured with the Dutch version of the Maslach Burnout Inventory (MBI) [25], which contains three subscales: (1) emotional exhaustion (MBI-EE), 5 items; (2) depersonalization (MBI-DP), 4 items; and (3) personal accomplishment (MBI-PA), 6 items. Each item is scored on a scale from 0 to 6, and subscale scores are calculated by adding the item scores and dividing this subscale total score by the number of items. For MBI-EE and MBI-DP, a higher score indicates more work-related stress, while a high MBI-PA score indicates less work-related stress. Individuals can be considered burnt out when they report high MBI-EE (≥ 2.2) in combination with high MBI-DP (≥ 2.0) or low MBI-PA (≤ 3.66) [26].

Quality of life was assessed with the EuroQoL questionnaire (EQ-5D) [27]. The EQ-5D consists of 5 items (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each of which is rated as causing "no problems," "some problems," or "extreme problems." The EQ-5D can thus describe 486 unique health states. Each of these health states has been empirically valued between 0 (poor health) and 1 (perfect health). The scores of our respondents were weighted with these values to derive a single summary index score.

Analyses

Missing Values

All analyses were performed on the intention-to-treat sample. Pre-test data were available for all participants. Missing values of post-test nonresponders (17%, 36/213) were handled by using multiple imputation procedure NORM [28] in statistical package R. In this procedure, missing data are imputed by regression analyses using available baseline data (demographics as well as data on baseline severity) from the responders as well as the nonresponders. This means that not every nonresponder received the same post-test score, but the post-test score was dependent on the particular characteristics as defined by baseline (eg, gender, age). This regression analyses was then repeated five times. The effectiveness analyses were then performed on each of the five resulting data files, and the five estimates were combined into a single overall estimate using the multiple imputation inference rules of Rubin [29]. This yielded proper *P* values and confidence intervals for the estimates. All reported *P* values are two-tailed.

Effectiveness

Effectiveness was calculated in three ways: (1) analyzing mean improvement scores, (2) calculating the proportion of participants who made significant improvements, and (3) calculating the proportion of participants who recovered. Each will be described in more detail below.

Mean Improvement Scores

The magnitude of the effect of the intervention (Cohen's *d*) was calculated by subtracting the post-test mean score of the control group (M_c) from the post-test mean score of the intervention group (M_i) and dividing the result by the pooled standard deviation (SD_{ic}). A Cohen's *d* of 0.5 thus indicates that the mean of the intervention group is half a standard deviation larger than the mean of the control group. Values of *d* from 0.56 to 1.2 can be assumed to be large, 0.33 to 0.55 are moderate, and 0 to 0.32 are small [30]. We calculated Cohen's *d* for all participants, participants who completed the intervention, and participants with severe baseline symptoms.

Significant Improvement

We calculated significant improvement as described by Jacobson and Truax [31]. We subtracted the pre-test score from the post-test score and divided the difference by its standard error. All participants falling below 1.96 (or above for MBI-PA and EQ-5D) were considered significantly improved since this amount of change is unlikely to occur by chance ($P < .05$). The differences in improvement rate between the intervention and control group were then calculated with binary logistic regression and expressed as odds ratios (ORs) and their 95% confidence intervals (95% CIs).

Recovery

A different definition of recovery was used for the different types of outcome. The definitions were as follows: (1) depression—no DSM-IV diagnoses of major depression according to the MDI, (2) anxiety—a HADS score lower than 8 (a score ≥ 8 is indicative of a general anxiety disorder [32], and (3) work-related stress—not meeting the burnout criteria of the MBI. This was calculated only for those participants who did meet these criteria at baseline. The differences in recovery rate between the intervention and control group were also calculated with binary logistic regression and expressed as odds ratios and their 95% confidence intervals.

Results

Response Rates

Out of 213 enrolled participants, 177 filled in the post-test questionnaires (response rate 83.1%). The response was significantly higher in the control group (91%; n = 96) than in the intervention group (76%, n = 81; $P = .004$). Furthermore, the response was higher among the more educated participants (94.9%; n = 111) than among less educated participants (69%, n = 66; $P < .001$) and higher among participants without alcohol problems (87.1%, n = 121) than among those with alcohol problems (76%, n = 56; $P = .04$).

All the baseline differences between responders and nonresponders on the outcome measures were in the same direction: nonresponders reported poorer health at baseline than responders. However, the differences were very small and not statistically significant (Table 1).

Table 1. Baseline scores of depression, anxiety, burnout, and quality of life (N = 213)

Scale	Responders (n = 177), Mean (SD)	Dropouts (n = 36), Mean (SD)	*P* Value
CES-D	29.8 (9.3)	30.2 (8.6)	.80
MDI	24.3 (9.1)	26.7 (10.2)	.16
SCL-A	23.8 (7.1)	24.7 (8.1)	.47
HADS	10.0 (3.2)	10.1 (3.6)	.93
MBI-EE	2.8 (1.4)	2.9 (1.3)	.76
MBI-PA	3.2 (1.0)	3.6 (1.2)	.12
MBI-DP	2.4 (1.4)	2.2 (1.4)	.60
EQ-5D	0.62 (0.23)	0.61 (0.25)	.81

Descriptive Analysis of Baseline Variables

As shown in Table 2, most participants in this study were female (71.4%; n = 152), born in the Netherlands (91.5%; n = 195), higher educated (54.9%; n = 117), and had a paid job (64.8%; n = 138). The participants in the intervention group were more often married (59.8%; n = 64) than participants in the control group (44.3%; n = 47; $P = .02$). There were no differences between the intervention and control groups with regard to baseline depression, anxiety, stress, or quality of life scores (Table 3).

Table 2. Baseline characteristics of the participants

Characteristic	All (N = 213), No. (%)	Intervention (N = 107), No. (%)	Control (N = 106), No. (%)	*P* Value
Gender				.85
Male	61 (28.6)	30(28.0)	31 (29.2)	
Female	152 (71.4)	77 (72.0)	75 (70.8)	
Married				.02
No	102 (47.9)	43 (40.2)	59 (55.7)	
Yes	111 (52.1)	64 (59.8)	47 (44.3)	
Country of birth				.31
Netherlands	195 (91.5)	100 (93.5)	95 (89.6)	
Other	18 (8.5)	7 (6.5)	11 (10.4)	
Education				.19
Lower	96 (45.1)	53 (49.5)	43 (40.6)	
Higher[*]	117 (54.9)	54 (50.5)	63 (59.4)	
Paid job				.85
No	75 (35.2)	37 (34.6)	38 (35.8)	
Yes	138 (64.8)	70 (65.4)	68 (64.2)	
Sick leave[†]				.32
No	111 (80.4)	54 (77.1)	57 (83.8)	
Yes	27 (19.6)	16 (22.9)	11 (16.2)	
Alcohol problems				.36
CAGE[‡] < 2	139 (65.3)	73 (68.2)	66 (62.3)	
CAGE ≥ 2	74 (34.7)	34 (31.8)	40 (37.7)	
Age, mean (SD)	45.2 (10.6)	45.1 (10.9)	45.4 (10.4)	.84

[*]Higher education equals higher vocational education or university.

[†]Calculated only for the 64.8% (n = 138) participants with a paid job.

[‡]The CAGE questionnaire is a screening test for alcohol dependence.

Adherence and Attrition

Of all 107 participants in the intervention group, 9% (n = 10) dropped out before the course started. The first assignment (Week 1) was completed by the remaining 91% (n = 97). Then another 17% (n = 18) dropped out, and the second assignment (Week 2) was completed by 74% (n = 79). Another 8% (n = 9) dropped out, and the third assignment (Week 3) was completed by 65% (n = 70). Finally, another 10% (n = 11) dropped out, leaving 55% (n = 59) who completed the whole course. Married participants more often completed the course (66%; n = 42) than non-married participants (40%, n = 17; *P* = .008). There were no other significant demographic or baseline differences between the participants who did or did not complete the course.

Mean Improvements Scores: Depression, Anxiety, Stress, and Quality of Life

In general, the intervention had a significant effect on symptoms of depression, anxiety, and quality of life but not on work-related stress (Table 3). The analyses of all participants showed the most profound effects for the CES-D (d = 0.50) and the SCL-A (d = 0.42). In general, the effect sizes were largest for those participants who fully completed the intervention (n = 59). For these, the intervention was most effective for depression (CES-D: d = 0.67; MDI: d = 0.56), but the results for anxiety (SCL-A: d = 0.51; HADS: d = 0.48) and quality of life (EQ-5D: d = 0.44) were also substantial.

In a subset analysis, we selected only the participants with the most severe problems at baseline and calculated their improvements for each measure (Table 4). Compared to all participants (see Table 3), those with the most severe problems at baseline improved more, as evidenced by higher effect sizes, with the exception of scores on the SCL-A scale, for which the effect size decreased from 0.42 to 0.37. Improvements in effect size were most notable for work-related stress: the overall effect size on the MBI-EE subscale was 0.28 for all participants but improved to 0.65 for participants who actually experienced a burnout at baseline.

Table 3. Effects of self-examination therapy on depression, anxiety, burnout, and quality of life

Scale*	Control (N = 106), Mean (SD)		Intervention, All (N = 107), Mean (SD)		Intervention, Course Completers (N = 59), Mean (SD)		Effect Size† (95% CI)	
	Pre-Test	Post-Test	Pre-Test	Post-Test	Pre-Test	Post-Test	All	Course Completers
CES-D	29.9 (9.2)	26.2 (10.5)	29.9 (9.1)	20.9 (10.8)	29.8 (8.5)	19.3 (10.1)	0.50 (0.22-0.79)	0.67 (0.32-1.02)
MDI	23.6 (9.0)	25.1 (6.8)	25.8 (9.6)	22.9 (6.9)	25.1 (8.9)	21.4 (6.2)	0.33 (0.03-0.63)	0.56 (0.22-0.90)
SCL-A	23.7 (7.2)	22.7 (7.5)	24.1 (7.4)	19.7 (6.8)	10.0 (2.9)	19.1 (6.2	0.42 (0.14-0.70)	0.51 (0.18-0.84)
HADS	9.9 (3.3)	9.1 (3.3)	10.1 (3.3)	8.0 (3.4)	24.2 (7.0)	7.5 (3.2)	0.33 (0.04-0.61)	0.48 (0.15-0.82)
MBI-EE	2.8 (1.5)	2.8 (1.5)	2.9 (1.3)	2.5 (1.5)	2.8 (1.1)	2.5 (1.4)	0.28 (−0.08 to 0.64)	0.20 (−0.26 to 0.66)
MBI-PA	3.4 (1.0)	3.2 (1.0)	3.2 (1.1)	3.5 (1.0)	2.2 (1.3)	3.5 (1.0)	0.33 (−0.03 to 0.69)	0.36 (−0.25 to 0.98)
MBI-DP	2.4 (1.4)	2.6 (1.5)	2.4 (1.3)	2.3 (1.4)	3.1 (1.2)	2.2 (1.5)	0.20 (−0.15 to 0.56)	0.27 (−0.22 to 0.75)
EQ-5D	0.61 (0.24)	0.66 (0.20)	0.62 (0.23)	0.73 (0.20)	0.63 (0.22)	0.8 (0.2)	0.31 (0.03-0.60)	0.44 (0.11-0.77)

*The values for the MBI subscales are only given for those with a paid job; n = 70 in the intervention condition; n = 68 in the control.

†Effect size is presented as Cohen's *d*: the number of standard deviations the intervention group has improved more than the control group; $(M_c - M_i) / Sd_{ic}$.

Table 4. Effects of self-examination therapy on the subset of participants with severe symptoms of depression, anxiety, burnout, and quality of life at baseline

Scale	Definition of Severe Symptoms	Control			Intervention			Effect Size* (95% CI)
		No.	Pre-Test, Mean (SD)	Post-Test, Mean (SD)	No.	Pre-Test, Mean (SD)	Post-Test, Mean (SD)	
CES-D	≥ 16	99	31.1 (8.1)	27.3 (9.8)	97	31.6 (7.6)	21.7 (10.8)	0.54 (0.25-0.84)
MDI	DSM-IV depression	37	32.8 (5.2)	28.3 (6.9)	44	33.7 (5.5)	25.5 (6.8)	0.41 (−0.04 to 0.86)
SCL-A	≥ 18	89	25.5 (6.4)	24.1 (7.3)	84	26.7 (6.0)	21.6 (6.4)	0.37 (0.06-0.69)
HADS	≥ 8	78	11.3 (2.5)	10.2 (3.0)	85	11.3 (2.6)	8.7 (3.3)	0.45 (0.13-0.78)
MBI-EE	burnout	34	3.9 (1.0)	3.8 (1.3)	43	3.4 (1.0)	2.9 (1.3)	0.65 (0.14-1.16)
MBI-PA	burnout	34	3.1 (0.9)	3.0 (0.9)	43	2.9 (1.0)	3.3 (1.1)	0.33 (−0.14 to 0.81)
MBI-DP	burnout	34	3.3 (1.1)	3.2 (1.4)	43	2.9 (1.3)	2.6 (1.5)	0.44 (−0.06 to 0.95)
EQ-5D	≥ 0.55	74	0.75 (0.06)	0.7 (0.2)	73	0.76 (0.08)	0.8 (0.2)	0.34 (0.00-0.69)

*Effect size is presented as Cohen's *d*: the number of standard deviations the intervention group has improved more than the control group; $(M_c - M_i) / Sd_{ic}$.

Significant Improvement

The proportion of participants with significant improvements (their change is so large it is unlikely to have occurred by chance, see definition under "Methods") in both groups is compared in Table 5. The results show significant effects of the intervention both for depression (CES-D and MDI) and anxiety (SCL-A and HADS). The improvements on the MBI-PA scale are also statistically significant. The differences between the intervention and the control groups for the remaining outcomes were all in favour of the intervention group (OR between 1.6 and 2.2), but these results were not statistically significant.

Table 5. Participants with significant improvement

Scale	Intervention (N = 107), No. (%)	Control (N = 106), No. (%)	OR	95% CI
CES-D	52 (48.4)	22 (20.9)	3.5	1.9-6.7
MDI	22 (20.7)	7 (6.6)	3.7	1.4-10.0
SCL-A	23 (21.3)	12 (11.3)	2.1	1.0-4.6
HADS	38 (35.9)	16 (15.5)	3.1	1.6-6.0
MBI-EE	14 (13.4)	7 (6.8)	2.2	0.6-8.1
MBI-PA	23 (21.4)	7 (6.5)	3.9	1.2-12.6
MBI-DP	8 (7.7)	5 (4.7)	1.7	0.4-7.1
EQ-5D	25 (23.7)	17 (16.2)	1.6	0.8-3.3

Recovery

Of all 81 participants who suffered major depression according to the MDI at baseline, a total of 52 (64.4%) had recovered at post-test across both groups (Table 6). Recovery occurred more often in the intervention group (72.7%) than in the control group (54.6%, OR = 2.2), but this effect was not statistically significant (95% CI 0.8-6.0). Recovery from anxiety and burnout also occurred more often in the intervention group than in the control group. However, the result with regard to anxiety was not statistically significant (OR = 2.0; 95% CI 0.9-4.2), while that for burnout was (OR = 4.0; 95% CI 1.2-13.5).

Table 6. Recovery of participants with depression, anxiety, and burnout (as established at baseline)

	Total No. Participants at Baseline	Definition of Recovery	Post-Test, No. (%)		OR	95% CI
			Intervention	Control		
Depression	81	No MDI diagnoses	32/44 (72.7)	20/37 (54.6)	2.2	0.8-6.0
Anxiety	134	HADS < 8	26/70 (37.7)	15/64 (23.4)	2.0	0.9-4.2
Burnout	77	No MBI diagnosis	16/43 (38.1)	5/34 (13.5)	4.0	1.2-13.5

Discussion

Principal Results

We studied the effects of a short, generic, Web-based, self-help intervention for mental health problems in a randomized trial among 213 participants with symptoms of depression, anxiety, or work-related stress. We demonstrated statistically and clinically significant effects on symptoms of depression and anxiety. These effects were even more pronounced among participants with more severe baseline problems and for participants who fully completed the course. The effects on work-related stress and quality of life were less clear.

Limitations

This study has several limitations. The first is related to the choice of the control group. We could have chosen a care-as-usual comparison (ie, not have given any intervention to the control group); however, this might have limited the generalizability of our results since in that case only patients willing to be randomized to a non-treatment option would have participated. It is likely that these patients differ from the ones who do want (need) treatment. We also might have chosen an attention placebo control group or comparison with another intervention. It is known that effects of attention placebo controlled trials are usually smaller than waitlist controlled trials. However, with our intervention, we especially intended to reach those people who do not get any treatment at all [33], and in this case, we feel that a comparison with a waiting list control group is justified. We stress that the demonstrated effects might, in part, be caused by common therapy factors and not by the specific intervention we studied (eg, the attention given to the intervention group by means of email support might have caused effects regardless of the contents of the feedback or the intervention). Nevertheless, since we intend to implement the course in the Netherlands as is (including support), this effect is what we wanted to measure.

The second limitation has to do with the response rate. Although the overall response rate was satisfactory (83%), the response rate of the intervention group was significantly lower (76%) than that of the control group (91%). We could find no indications for selection bias since we could not demonstrate clear baseline differences between the responders and nonresponders (except for marital status). The bias that still might have been introduced was accounted for by imputing all missing data (multiple imputations) and performing intention-to-treat analyses. Nevertheless, imputing 24% of the data might have led to unreliable estimates.

Another limitation is the fact that participants could only be included in the study if they had computer skills and access to Internet. Thus, the participants in this study were more highly educated than the general population, and it is uncertain whether the results of this study can be generalized to people with less education.

Specific Findings

Meta-analyses for bibliotherapy regarding different types of target problems have shown effect sizes between 0.53 and 0.96 [3]. For depression and anxiety, a recent meta-analysis of Web-based, cognitive behavioral, self-help interventions showed mean effect sizes (Cohen's *d*) of 0.32 and 0.96, respectively [5]. Thus, our results on symptoms of depression and anxiety seem to fit well within the reported range. It is important to note that our results were obtained in less time (4 weeks) than is usual for Web-based interventions for anxiety or depression (often 6 weeks or more). Furthermore, our results are also almost identical to those found in a meta-analysis of face-to-face problem-solving treatment ($d = 0.42$) [34]. All this implies that our intervention may be a worthwhile alternative to other more intensive or expensive treatment options, especially since it can be used for participants with comorbid symptoms of anxiety and depression. However, longer follow-up studies are necessary to determine the treatment gains over a longer period of time.

The results with regard to work-related stress were less consistent. When considering only those participants who were suffering from burnout at the start of the study, the results were promising. The participants in the intervention group were four times (95% CI 1.2-13.5) more likely to recover from their burnout than participants in the control group, and they experienced a substantial improvement with regard to the EE subscale of the MBI (Cohen's $d = 0.65$). These effects disappeared when considering all participants (or all participants who completed the intervention). This probably can be explained by the relatively small percentage of participants who actually did experience work-related stress at the start of the study: only 77 participants (36%) could be described as suffering from burnout. Furthermore, it must be noted that, in general, the effects of interventions for work-related stress seem to be less pronounced. Meta-analyses have reported effect sizes between 0.35 and 0.68 for different types of face-to-face intervention [35].

Conclusion

To our knowledge, this is the first trial on a short, Web-based, problem-solving intervention for participants with different types of (comorbid) emotional problems. The results seem to be as good as other longer, disease-specific bibliotherapies. Longitudinal research is needed to study the long-term effects.

Acknowledgments

The study was performed at the Vrije Universiteit Amsterdam, the Netherlands, without additional financial support.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Intervention screenshots

[PDF file (Adobe PDF), 304 KB - jmir_v10i1e7_app1.pdf]

References

1. Cuijpers P. Bibliotherapy in unipolar depression: a meta-analysis. J Behav Ther Exp Psychiatry 1997 Jun;28(2):139-147. [Medline: 9194011] [doi: 10.1016/S0005-7916(97)00005-0]
2. Apodaca TR, Miller WR. A meta-analysis of the effectiveness of bibliotherapy for alcohol problems. J Clin Psychol 2003 Mar;59(3):289-304. [Medline: 12579546] [doi: 10.1002/jclp.10130]
3. den Boer PCAM, Wiersma D, Van den Bosch RJ. Why is self-help neglected in the treatment of emotional disorders? A meta-analysis. Psychol Med 2004 Aug;34(6):959-971. [Medline: 15554567]
4. Anderson L, Lewis G, Araya R, Elgie R, Harrison G, Proudfoot J, et al. Self-help books for depression: how can practitioners and patients make the right choice? Br J Gen Pract 2005 May;55(514):387-392 [FREE Full text] [Medline: 15904559]
5. Spek V, Cuijpers P, Nyklíĉek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behavior therapy for mood and anxiety disorders: a meta-analysis. Psychol Med 2007;37.
6. Marrs RW. A meta-analysis of bibliotherapy studies. Am J Community Psychol 1995 Dec;23(6):843-870. [Medline: 8638553] [doi: 10.1007/BF02507018]
7. Clarke G, Reid E, Eubanks D, O'Connor E, DeBar LL, Kelleher C, et al. Overcoming depression on the Internet (ODIN): a randomized controlled trial of an Internet depression skills intervention program. J Med Internet Res 2002 Dec 17;4(3):E14 [FREE Full text] [Medline: 12554545] [doi: 10.2196/jmir.4.3.e14]
8. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [Medline: 14742346] [doi: 10.1136/bmj.37945.566632.EE]
9. Bijl RV, Ravelli A. Psychiatric morbidity, service use, and need for care in the general population: results of The Netherlands Mental Health Survey and Incidence Study. Am J Public Health 2000 Apr;90(4):602-607 [FREE Full text] [Medline: 10754976]

10. Andrews G, Henderson S, Hall W. Prevalence, comorbidity, disability and service utilisation. Overview of the Australian National Mental Health Survey. Br J Psychiatry 2001 Feb;178(2):145-153 [FREE Full text] [Medline: 11157427] [doi: 10.1192/bjp.178.2.145]
11. Nezu AM. Efficacy of a social problem-solving therapy approach for unipolar depression. J Consult Clin Psychol 1986 Apr;54(2):196-202. [Medline: 3700806] [doi: 10.1037/0022-006X.54.2.196]
12. Mynors-Wallis LM, Gath DH, Day A, Baker F. Randomised controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care. BMJ 2000 Jan 1;320(7226):26-30 [FREE Full text] [Medline: 10617523] [doi: 10.1136/bmj.320.7226.26]
13. Bowman D, Scogin F, Floyd M. Efficacy of self-examination therapy in the treatment of generalized anxiety disorder. J Couns Psychol 1997;44(3). [doi: 10.1037/0022-0167.44.3.267]
14. Floyd M, McKendree-Smith N, Bailey E, Stump J, Scogin F, Bowman D. Two-year follow-up of self-examination therapy for generalized anxiety disorder. J Anxiety Disord 2002;16(4):369-375. [Medline: 12213033] [doi: 10.1016/S0887-6185(02)00105-6]
15. van der Klink JJL, Blonk RWB, Schene AH, van Dijk FJH. Reducing long term sickness absence by an activating intervention in adjustment disorders: a cluster randomised controlled design. Occup Environ Med 2003 Jun;60(6):429-437 [FREE Full text] [Medline: 12771395]
16. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. J Med Internet Res 2002 Feb 15;4(1):e3 [FREE Full text] [Medline: 11956035] [doi: 10.2196/jmir.4.1.e3]
17. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [Medline: 14742346] [doi: 10.1136/bmj.37945.566632.EE]
18. Bowman DG. Self-examination therapy: treatment for anxiety and depression. In: VandeCreek L, editor. Innovations in Clinical Practice: A Source Book. Sarasota, FL: Professional Resource Press/Professional Resource Exchange; 1995.
19. Bowman V, Ward LC, Bowman D, Scogin F. Self-examination therapy as an adjunct treatment for depressive symptoms in substance abusing patients. Addict Behav 1996;21(1):129-133. [Medline: 8729714] [doi: 10.1016/0306-4603(95)00027-5]
20. Bowman D, Scogin F, Lyrene B. The efficacy of self-examination therapy and cognitive bibliotherapy in the treatment of mild to moderate depression. Psychotherapy Research 1995;5(2):131-140.
21. Radloff L. The CES-D scale: a self report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
22. Bech P, Rasmussen NA, Olsen LR, Noerholm V, Abildgaard W. The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination as the index of diagnostic validity. J Affect Disord 2001 Oct;66(2-3):159-164. [Medline: 11578668] [doi: 10.1016/S0165-0327(00)00309-8]
23. 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] [doi: 10.1017/S0033291796004382]
24. Derogatis LR. SCL-90: Administrative, Scoring, and Procedures Manual-I for the Revised Version. Baltimore, MD: Johns Hopkins University School of Medicine, Clinical Psychometrics Research Unit; 1977.
25. Maslach C, Schaufeli WB, Leiter MP. Job burnout. Annu Rev Psychol 2001;52(1):397-422. [Medline: 11148311] [doi: 10.1146/annurev.psych.52.1.397]
26. Brenninkmeijer V, VanYperen N. How to conduct research on burnout: advantages and disadvantages of a unidimensional approach in burnout research. Occup Environ Med 2003 Jun;60 Suppl 1:i16-i20 [FREE Full text] [Medline: 12782742] [doi: 10.1136/oem.60.suppl_1.i16]
27. Brooks R. EuroQol: the current state of play. Health Policy 1996 Jul;37(1):53-72. [Medline: 10158943] [doi: 10.1016/0168-8510(96)00822-6]
28. Schafer JL. Analysis of Incomplete Multivariate Data. London, UK: Chapman and Hall; 1997.
29. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.
30. Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. Am Psychol 1993 Dec;48(12):1181-1209. [Medline: 8297057] [doi: 10.1037/0003-066X.48.12.1181]
31. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [Medline: 2002127] [doi: 10.1037/0022-006X.59.1.12]
32. Olssøn I, Mykletun A, Dahl AA. The Hospital Anxiety and Depression Rating Scale: a cross-sectional study of psychometrics and case finding abilities in general practice. BMC Psychiatry 2005;5(1):46 [FREE Full text] [Medline: 16351733] [doi: 10.1186/1471-244X-5-46]
33. Wang PS, Aguilar-Gaxiola S, Alonso J, Angermeyer MC, Borges G, Bromet EJ, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet 2007 Sep 8;370(9590):841-850. [Medline: 17826169] [doi: 10.1016/S0140-6736(07)61414-7]
34. Cuijpers P, van Straten A, Warmerdam L. Problem solving therapies for depression: a meta-analysis. Eur Psychiatry 2007 Jan;22(1):9-15. [Medline: 17194572] [doi: 10.1016/j.eurpsy.2006.11.001]

35. van der Klink JJ, Blonk RW, Schene AH, van Dijk FJ. The benefits of interventions for work-related stress. Am J Public Health 2001 Feb;91(2):270-276 [FREE Full text] [Medline: 11211637]

Abbreviations

CES-D: Center for Epidemiological Studies Depression Scale
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
EQ-5D: EuroQoL questionnaire
HADS: Hospital Anxiety and Depression Scale (only Anxiety section is used)
MDI: Major Depression Inventory
MBI: Maslach Burnout Inventory
MBI-DP: MBI Depersonalization subscale
MBI-EE: MBI Emotional Exhaustion subscale
MBI-PA: MBI Personal Accomplishment subscale
SCL-A: Symptom Checklist – Anxiety section

submitted 10.07.07; peer-reviewed by R Munoz, K Griffiths, B Danaher; comments to author 23.10.07; revised version received 20.12.07; accepted 04.01.08; published 25.03.08.

Please cite as:
van Straten A, Cuijpers P, Smits N
Effectiveness of a Web-Based Self-Help Intervention for Symptoms of Depression, Anxiety, and Stress: Randomized Controlled Trial
J Med Internet Res 2008;10(1):e7
URL: http://www.jmir.org/2008/1/e7/
doi:10.2196/jmir.954
PMID:18364344

Original Paper

Web Evaluation at the US National Institutes of Health: Use of the American Customer Satisfaction Index Online Customer Survey

Fred B Wood[1], MBA, DBA; Elliot R Siegel[1], PhD; Sue Feldman[2]; Cynthia B Love[1], MLS; Dennis Rodrigues[3]; Mark Malamud[4]; Marie Lagana[5], MA; Jennifer Crafts[6], PhD

[1]National Library of Medicine, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD, USA

[2]National Cancer Institute, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD, USA

[3]Office of Communications and Public Liaison, Office of the Director, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD, USA

[4]National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD, USA

[5]Center for Information Technology, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD, USA

[6]Westat Inc, Rockville, MD, USA

Corresponding Author:
Fred B Wood, MBA, DBA
US National Institutes of Health
US Department of Health and Human Services
Office of Health Information Programs Development
US National Library of Medicine
8600 Rockville Pike, Bldg 38, Room 2S-14
Bethesda, MD 20894
USA
Phone: +1 301 402 9278
Fax: +1 301 496 4450
Email: fredwood@mail.nih.gov

Abstract

Background: The National Institutes of Health (NIH), US Department of Health and Human Services (HHS), realized the need to better understand its Web users in order to help assure that websites are user friendly and well designed for effective information dissemination. A trans-NIH group proposed a trans-NIH project to implement an online customer survey, known as the American Customer Satisfaction Index (ACSI) survey, on a large number of NIH websites—the first "enterprise-wide" ACSI application, and probably the largest enterprise Web evaluation of any kind, in the US government. The proposal was funded by the NIH Evaluation Set-Aside Program for two years at a cost of US $1.5 million (US $1.275 million for survey licenses for 60 websites at US $18,000 per website; US $225,000 for a project evaluation contractor).

Objective: The overall project objectives were to assess the value added to the participating NIH websites of using the ACSI online survey, identify any NIH-wide benefits (and limitations) of the ACSI, ascertain any new understanding about the NIH Web presence based on ACSI survey results, and evaluate the effectiveness of a trans-NIH approach to Web evaluation. This was not an experimental study and was not intended to evaluate the ACSI survey methodology, per se, or the impacts of its use on customer satisfaction with NIH websites.

Methods: The evaluation methodology included baseline pre-project websites profiles; before and after email surveys of participating website teams; interviews with a representative cross-section of website staff; observations of debriefing meetings with website teams; observations at quarterly trans-NIH Web staff meetings and biweekly trans-NIH leadership team meetings; and review and analysis of secondary data.

Results: Of the original 60 NIH websites signed up, 55 implemented the ACSI survey, 42 generated sufficient data for formal reporting of survey results for their sites, and 51 completed the final project survey. A broad cross-section of websites participated, and a majority reported significant benefits and new knowledge gained from the ACSI survey results. NIH websites as a group scored consistently higher on overall customer satisfaction relative to US government-wide and private sector benchmarks.

Conclusions: Overall, the enterprise-wide experiment was successful. On the level of individual websites, the project confirmed the value of online customer surveys as a Web evaluation method. The evaluation results indicated that successful use of the ACSI, whether site-by-site or enterprise-wide, depends in large part on strong staff and management support and adequate funding and time for the use of such evaluative methods. In the age of Web-based e-government, a broad commitment to Web evaluation may well be needed. This commitment would help assure that the potential of the Web and other information technologies to improve customer and citizen satisfaction is fully realized.

(J Med Internet Res 2008;10(1):e4) doi:10.2196/jmir.944

KEYWORDS

Surveys; evaluation studies; satisfaction; Internet; World Wide Web; consumer health information

Introduction

At the US National Institutes of Health (NIH), as at many other biomedical institutions, World Wide Web–based information dissemination now dominates [1,2]. The use of the Internet and Web at NIH has grown dramatically over the last decade, to the point where all major NIH organizations have one or more websites. NIH has realized the necessity to better understand Web users in order to help assure that websites are user friendly and well designed for effective information dissemination.

Multidimensional Approach

Over the last several years, various individual NIH organizations have experimented with several different methods of Web evaluation [3-5]. These methods have evolved into a so-called "multidimensional approach" to Web evaluation that acknowledges that no one evaluation method meets all needs. Methods may vary with the preferences and sophistication of individual website teams, complexity of websites, and stage of the website improvement cycle.

The multidimensional approach can be described as including methods in four categories: usability testing, user feedback, usage data, and website and Internet performance data. These methods are primarily based on feedback from both users and the systems that monitor Web servers and Internet performance [6].

Another way to describe the multidimensional approach divides evaluation methods into two groups: what users say about a website, and what experts say. Prior to the American Customer Satisfaction Index (ACSI) project reported here, NIH as a whole placed the greater emphasis on evaluating its website content by "what experts say," ensuring quality information through writing and review of Web content by subject experts. This ACSI project is one step in giving Web teams at NIH another tool to learn more about "what users say."

User opinions and behavior—what users say—are expressed through Web logs, surveys, focus groups, email, phone, personal contact, words used in search queries, Internet audience measurement, usability studies, and other methods [6,7]. NIH websites vary considerably in the budget, staff, and time they have to implement Web evaluation based on user input. For many of the websites participating in this study, this ACSI project was their first opportunity to get routine, structured, direct-from-the-user feedback.

Customer satisfaction surveys, like the ACSI, are one tool for listening to "what users say" to determine user perceptions of a website's usefulness and performance. Perceptions are inherently subjective, but they do help Web managers understand another facet of user opinion. Other prior user-based evaluations at NIH have included search log analysis of user queries on a website or user queries on referring sites such as major Internet portals [8,9], analysis of email from users [10], and research on market share for online health information services [11].

The second group of Web evaluation methods, what experts say, is already heavily used in evaluating NIH Web content because of the inherent importance of providing accurate health information that can be accessed by many different audiences. NIH organizations have focused considerable efforts on ensuring that their websites convey the highest quality health information and reflect the latest findings from medical research.

Especially in health and medicine, subjectively perceived customer satisfaction can be only one measure of the value of a health information website. Some users might readily find a well-designed website with convincing graphics, testimonials, and popular appeal to be "highly satisfactory," even if the site's health information content is misleading, erroneous, or even harmful. NIH websites aim to be both well-designed and credible. Examples of NIH's strengths in evaluating content include the efforts of websites to use strict guidelines for selecting and writing health content [12,13], evaluate content for readability and ethnic/cultural sensitivity [14], fund and implement research on Web design of health information for children, seniors [15], and others, and to secure external accreditation from organizations like Health On the Net (HON) Foundation [16].

In the age of Web-based e-government, a broad commitment to Web evaluation may well be needed. This commitment would help assure that the potential of the Web and other information technologies to improve customer and citizen satisfaction is fully realized.

Customer Satisfaction

In parallel with the rise to dominance of the Internet and Web has been an increasing emphasis on "customer satisfaction" in the US government. Customer satisfaction is viewed as an important metric of the political goal of developing a more "customer-centric" government that is more responsive to citizen needs. These needs include a wide range of types of information from the government. In the case of NIH, citizens are seeking

biomedical and health information on diverse diseases, conditions, health trends, research results, and the like.

There are many examples of requirements for the federal government to address customer needs and satisfaction. The Government Performance and Results Act of 1993 states the following: "The purposes of this act are to…improve Federal program effectiveness and public accountability by promoting a new focus on results, service quality, and *customer satisfaction*" [17] (italics added for emphasis).

The Office of Management and Budget (OMB) Circular A-130: Management of Federal Information Resources requires agencies to develop enterprise architecture that "will define principles and goals and set direction on such issues as the promotion of interoperability, open systems, public access, compliance with GPRA, *end user satisfaction*, and IT security" [18] (italics added for emphasis). The OMB Circular also requires demonstrating "a projected return on the investment that is clearly equal to or better than alternative uses of available public resources. The return may include improved mission performance in accordance with Government Performance and Results Act measures, reduced cost, increased quality, speed, or flexibility; as well as *increased customer and employee satisfaction* [18, 19] (italics added for emphasis).

In 2004, the Interagency Committee on Government Information wrote "Recommended Policies and Guidelines for Federal Public Websites" [20] at the request of OMB. The suggestions formed the basis for "Policies for Federal Agency Public Websites" [21] issued by OMB. The recommended policies document includes extensive implementation guidance, currently used by federal Web managers, and suggests the use of "customer satisfaction surveys." Of key importance to the use of the ACSI at NIH is this provision:

> ***2e. Requirement: Organizations Must Measure Customer Satisfaction and Usability of Federal Public websites.*** *Organizations must evaluate customer satisfaction and usability of their websites and use the assessments to improve the websites. Federal public websites that reach the widest audiences—including agency websites and all second-level domain names registered in .gov, .mil, or .fed.us—must use a standard customer satisfaction survey.*
>
> *Rationale: Organizations that create federal public websites, and the citizens they serve, want these websites to be as useful as possible. While Web content managers do their best to write and organize their websites to be effective, they need to test their websites to identify problem areas and then fix those problems. A common customer satisfaction survey will reduce costs government-wide and compare government websites with each other.*

Online User Survey

Within the multidimensional evaluative approach, the online user survey is the method that provides the most direct feedback from users. Online user surveys can generate data on the types of users coming to a website, user demographics, levels of user satisfaction with the website and the information provided, and intended use of the information obtained.

Various NIH organizations, and in particular the National Library of Medicine (NLM) and National Cancer Institute (NCI), have a long history with user surveys, dating from the pre-Web era. NLM, for example, transitioned from paper to online surveys in the early 1980s and then to Web-based surveys in the late 1990s. These were snapshot surveys—typically fielded for 2 or 3 weeks—and only provided a "snapshot" of the customer base and were implemented at most once a year [22,23, personal communication, Cindy Love, National Library of Medicine, April 30, 2007]. In addition, there were few standard methods or benchmarks for surveys of websites.

In comparison, the ACSI methodology offers several advantages: continuous data collection, randomized rolling sample, rigorous standardized survey methodology, standardized questions plus capability for optional custom questions, and extensive benchmarking of results.

The ACSI was first implemented in 1994 as an offline survey measuring customer satisfaction with businesses [24] and was adapted to the Internet in 2002 [25]. More than two dozen other federal websites began using the survey in 2002 [26].

During the late 1990s, the President's Management Council, composed of the chief operating officers of each cabinet-level agency, responded to then Vice President Al Gore's National Performance Review (also known as the National Partnership for Reinventing Government) initiative by considering ways to measure citizen satisfaction with government services. The Council members and other government leaders were interested in measuring government services using the same methods as the private sector and holding government programs to a level of customer responsiveness equal to or better than the private sector [27]. The Council, with the Government Services Administration (GSA) taking the lead, solicited proposals for a measurement tool that could be used across multiple agencies and provide benchmarking among agencies and between government and nongovernment providers of services or goods.

In 1999, using federal contract competition processes, GSA awarded the contract to Arthur Andersen LLC and the University of Michigan to provide the ACSI for wide adoption as a survey measure of offline government services [28]. The ACSI was already well established as a measure of customer satisfaction in nongovernment sectors, routinely publishing its results in the Wall Street Journal and other prominent publications. This was the first opportunity for government agencies to use the same yardstick. GSA successfully sought clearance under the Paperwork Reduction Act for blanket permission for any agency to use the survey.

The contracting function and survey clearance responsibilities were assumed by the Federal Consulting Group in January 2000. The ACSI serves a unique role as the most widely and easily available survey instrument for federal government. Early users of the offline ACSI included the agencies that have the greatest contact with citizens such as the Social Security Administration (retirement beneficiaries), the Internal Revenue Service (tax filers), the State Department (passport applicants), the Customs

Service (international travelers), the Department of Veterans Affairs (compensation and medical care beneficiaries), and others. In October 2001, the ACSI also became available for online use through contract arrangements between ForeSee Results, Inc. and the Federal Consulting Group. The first online use was piloted by GSA for firstgov.gov (now USA.gov) and by NASA for NASA.gov. By mid-2002, the Federal Consulting Group obtained a generic clearance from the Office of Management and Budget for agencies that used the ForeSee Results Web metric tool and began to promote the use of the ACSI to federal Web managers. The ACSI continues to be used government-wide for both online and offline measures of customer satisfaction (personal communication by Bernie Lubran, ForeSee Results, Inc., May 1, 2007). Aggregate results for all government use of the ACSI, offline and online, are released every December by the University of Michigan [29-31].

NLM and NCI implemented the ACSI on several websites in 2003, taking advantage of the newly available contract providing the ACSI for measuring federal websites. In 2004, NLM and NCI staff shared their ACSI experience and survey results with the broader NIH Web community. This community, represented by a group known as the NIH Web Authors Group, was polled about their interest in participating in a trans-NIH project using the ACSI as a common online survey method.

The Web Authors Group members indicated strong interest, and as a result, a team of co-principal investigators self-organized to develop an evaluation plan and funding proposal. In mid-2004, a proposal was submitted to the NIH Evaluation Set-Aside Program and was approved for funding beginning in September 2004. The NIH Evaluation Branch [32] administers the Evaluation Set-Aside Program [33] that provides funds to evaluate programs and services at NIH. The US Department of Health and Human Services (HHS) “sets aside” funds each year for evaluation; institutes can then competitively apply for those funds. For NIH Web services, the Evaluation Branch funds several types of evaluation, depending on applications received. These have included feasibility studies, surveys (ACSI and others), usability, focus groups, user interviews, and measures of Internet connectivity.

The project was noteworthy because it was the first time that a broad cross-section of NIH organizations used the same method to evaluate websites. The implementation of website evaluations, as well as an external evaluation of the project, was designed and coordinated by a trans-NIH team of senior professionals. At peak participation, the project included 18 (of 27) NIH institutes and centers and 13 offices of the Office of the NIH Director, and 60 separate ACSI website licenses. See Table 1 for a list of participating NIH organizations and Multimedia Appendix 1 (Appendix A) for a list of specific websites.

Table 1. NIH organizations participating in the trans-NIH ACSI project (See Multimedia Appendix 1 [Appendix A] for a list of the specific websites participating in the project).

Institute/Center/Office	No. of ACSI Licenses
Institute/Center	
National Cancer Institute	7
National Eye Institute	1
National Human Genome Research Institute	1
National Heart, Lung, and Blood Institute*	6 (5)†
National Institute of Allergy and Infectious Diseases	1
National Institute of Arthritis and Musculoskeletal and Skin Diseases	1
National Institute on Drug Abuse	2
National Institute on Deafness and Other Communication Disorders	3
National Institute of Dental and Craniofacial Research	1
National Institute of General Medical Sciences	1
National Institute of Mental Health	1
National Library of Medicine	7
Center for Information Technology*	7 (3)†
National Center for Complementary and Alternative Medicine	1
Fogarty International Center	1
National Institute on Aging	1
National Institute of Diabetes and Digestive and Kidney Diseases	1
National Institute of Environmental Health Sciences	1
Total = 18	44 (39)†
Offices Within the NIH Office of the Director (OD)	
Office of Animal Care and Use	1
Office of Communications and Public Liaison	2
Office of Extramural Research	2
Office of Electronic Research and Reports Management	1
Office of Human Resources	1
Office of Research Services	1
Office of Research Facilities	1
Office of Rare Diseases	2
Office of Intramural Research Continuing Medical Education	1
Office of Dietary Supplements	1
Office of Technology Transfer	1
Office of Science Policy/Office of Science Education	2
Office of Science Policy and Planning*	1
Total = 13	17

*These NIH institutes and centers reallocated licenses to other websites or absorbed some license months into existing active licenses.

†Number of ACSI licenses allocated, with actual number of licenses used in parentheses.

The trans-NIH ACSI evaluation project lasted for two years, from September 2004 until September 2006, with initial and supplemental funding totaling US $1.5 million from the NIH Evaluation Set-Aside Program. This funding was for outside contracting of the ACSI survey implementation, offered by ForeSee Results Inc. [34], through the Federal Consulting Group / US Department of the Treasury [35], and for an outside evaluation conducted by Westat, Inc. The ACSI survey licenses

cost US $18,000 per website, for a total of US $1.275 million (the US $18,000 per site was considered competitive or less expensive for the value added compared to other survey options). The overall project evaluation by Westat, Inc. cost US $225,000. The contractors worked closely with the NIH co-principal investigators and leadership team and the participating NIH organizations.

This paper presents the results of the overall project evaluation that was concluded in fall 2006.

Methods

The core purpose of the project evaluation was to assess the value of using the ACSI to the participating NIH organizations, identify any NIH-wide benefits of the ACSI, ascertain any additional or new understanding about the NIH Web presence resulting from the ACSI, and evaluate the process of implementing an enterprise-wide approach.

It is important to note that the purpose was not to evaluate the ACSI itself as a stand-alone online survey methodology and/or as compared to other Web evaluation methods. The emphasis in this study was on the process of trans-NIH collaboration on Web evaluation, which was and still is unprecedented in scale. The ability to do an experimental study was confounded in part because websites started and ended their participation at variable times and because many websites did not participate long enough to go through a complete redesign cycle. Also, the emphasis of the study was not to increase ACSI customer satisfaction scores per se but to increase the familiarity of Web teams with use of online surveys as part of website evaluation. Finally, as will be noted in the discussion, the actual change in measured ACSI satisfaction scores when available was, in most cases, not statistically significant. For all these reasons, this project is properly viewed as an observational process study and not an experimental study.

The ACSI Methodology

The core ACSI methodology was developed by Professor Claes Fornell, Director of the National Quality Research Center, University of Michigan Business School, and is offered as an online service by ForeSee Results, Inc. of Ann Arbor, Michigan [36,37]. The ACSI method uses multiple regression analysis to link questions on key elements driving customer satisfaction with questions on overall customer satisfaction that are in turn linked to questions on future customer behavior. All standardized questions are framed using a 10-point Likert scale. The standardized questions cover the following areas: Elements that Drive Customer Satisfaction (ie, questions covering content, functionality, image, look and feel, navigation, search, privacy, and site performance); Composite Satisfaction (three questions); Future Behavior (ie, three questions covering likelihood to return, likelihood to recommend, likelihood to use as a primary resource).

Table 2 provides a complete list of the standardized ACSI questions. See Multimedia Appendix 2 for illustrations of the ACSI data reporting structure and analytical framework.

Table 2. Standardized questions used in the American Customer Satisfaction Index (ACSI) survey methodology

Category	Question*
	Please rate the following on a 10-point Likert scale.
	1 = poor, 10 = excellent
Site Performance	Speed of loading the page on this site?
	Consistency of speed on this site?
	Reliability of site performance on this site?
Search	Usefulness of search results on this site?
	Provides comprehensive search results on this site?
	Organization of search results on this site?
	Search features help you narrow the results on this site?
Privacy	Ability to limit sharing of your personal information on this site?
	Amount of personal information you are asked to submit on this site?
	Site's commitment to protecting your personal information?
Navigation	Number of steps to get where you want on this site?
	Ability to find information you want on this site?
	Clarity of site map or directory?
	Ease of navigation on this site?
Look and Feel	Ease of reading this site?
	Clarity of site organization?
	Clean layout on this site?
Functionality	Usefulness of the information provided on this site?
	Convenience of the information on this site?
	Ability to accomplish what you wanted to on this site?
Content	Accuracy of information on this site?
	Quality of information on this site?
	Freshness of content on this site?
	1 = very low, 10 = very high
Satisfaction	What is your overall satisfaction with this site?
	How well does this site meet your expectations?
	How does this site compare to your idea of an ideal Website?
	1 = very unlikely, 10 = very likely
Primary Resource	How likely are you to use this site as your primary resource for health information?
Recommend	How likely are you to recommend this site to someone else?
Likelihood to Return	How likely are you to return to this site?

*These standardized questions are taken from the ACSI online customer survey as used in this study.

In addition to standardized questions, the ACSI methodology allows for the inclusion of questions customized to specific client needs. Custom questions can have flexible formats, ranging from multiple-choice to open-ended.

Typical custom questions used in the NIH project included topics such as frequency of visits (eg, daily, weekly, monthly, first time); customer role (consumer, health provider, researcher, etc); primary purpose for visiting the website; primary means of finding the site; type of information being sought; demographics (age, gender, race/ethnicity, etc); results of query or search; use of the information found; and open-ended questions focusing on a site's strengths and weaknesses.

The ACSI survey used randomized selection with pop-up presentation of the survey. The sampling rate is set as a function of website traffic volume and estimated response rate, in order to obtain about 300 complete responses per 6-week reporting period. The typical response rate for participating NIH websites was about 5% (range of about 3% to 7%), and the sampling rate varied between a few percent (or less, the lowest being 0.1%) for the busiest sites to 100% for the low-traffic sites. The ACSI, like all online survey methods, can be problematic for very low traffic sites (see later discussion).

The GSA selected the ACSI in 1999 through a competitive procurement process for use by any interested government agencies. The Federal Consulting Group of the US Department of the Treasury now coordinates the government's contract with

ForeSee Results and interagency agreements between the Federal Consulting Group and agencies using the survey. The Federal Consulting Group also secures multi-year approval from OMB for the use of the ACSI survey by any federal agency. Under the requirements of the Paperwork Reduction Act of 1995, OMB must approve each collection of information by a federal agency (including customer satisfaction surveys) before it can be implemented. As part of its approval of the ACSI, OMB also provides expedited clearance of custom questions that are submitted in conjunction with the ACSI. If the OMB clearance through the Federal Consulting Group were not in place, each agency would need to allow several months to obtain the same clearance for each survey. By handling contracts and coordinating OMB clearances, the Federal Consulting Group greatly streamlines the process of survey implementation for participating federal agencies such as NIH.

Evaluation Methodology

The major evaluation component of the trans-NIH ACSI project was, in effect, an "evaluation of the evaluation," with greatest emphasis on the overall impact and utility of the ACSI at the website, organizational, and trans-NIH levels. Of the total project contracting budget of US $1.5 million, about US $225,000 was allocated to evaluation.

The evaluation contractor, Westat, Inc., was engaged throughout the project, worked closely with the NIH leadership team, and attended quarterly trans-NIH meetings with staff from participating websites.

The major components of the project evaluation strategy included the following. At the outset of the study, baseline website profiles were completed for all sites participating in the evaluation. These profiles were established in order to provide a baseline understanding of each site. The profiles were based on self-reported measures by website teams and coding of site characteristics (including website purpose, users, traffic levels, etc).

At the beginning and end of the study, email surveys of participating website teams were conducted. A total of 51 websites completed both the before and after surveys. The response rate for the final Web team survey was 51 out of 55 that implemented the ACSI, or 93%. Also at the beginning and end, the evaluation contractor interviewed a representative cross-section of website staff. Staff from about one third of the websites were interviewed one or more times. Teams were selected for interviewing so as to be representative of website size, purpose, and experience using the ACSI.

During the course of the study, ForeSee Results debriefing meetings with website teams were observed by the evaluation contractor. ForeSee Results, the ACSI contractor, held quarterly meetings, mostly by teleconference, with participating Web teams to discuss survey results and analysis. The NIH evaluation contractor observed a cross-section of these meetings. The evaluation contractor also observed discussions at quarterly trans-NIH ACSI meetings. The trans-NIH leadership team convened quarterly meetings for participating NIH staff to discuss progress, interim results, and lessons learned. ForeSee Results, Westat, and the Federal Consulting Group typically attended these quarterly meetings and gave brief presentations, fielded questions, and engaged in discussion as appropriate. The evaluation contractor also observed discussions at biweekly meetings of the trans-NIH leadership team.

Finally, in addition to the primary data collection listed above, the evaluation contractor had the benefit of secondary data, including quarterly reports on government-wide and private sector ACSI customer satisfaction results. These data were used to track performance of NIH websites and benchmark them against government and private sector websites with similar functions. Multimedia Appendix 3 includes all ACSI quarterly reports on overall ACSI survey results, from inception through March 2007, for federal agencies participating in the e-government satisfaction index based on the ACSI.

The completeness and robustness of the overall project evaluation strategy is illustrated in Table 3 and by specific website in the matrix included in Multimedia Appendix 1 (Appendix A). Multimedia Appendix 1 (Appendix B) also includes copies of the initial and final website staff survey instruments and the initial and final website staff interview instruments.

Table 3. Evaluation methods and data sources for the trans-NIH ACSI project

Method/Data Source	Primary Content	Planned Coverage (actual n)
Review of secondary data		
Website review	• Coding of a variety of website characteristics	All sites (61)
ForeSee pre-implementation worksheets	• Coding of team's responses to pre-implementation questions	All sites (48)
ACSI data for sites generating sufficient response for model data	• Satisfaction results per quarter	All sites collecting data during evaluation period: Q4 2004 (8) through Q1 2006 (42)
ACSI site-level data aggregated to NIH level	• Standard custom question results • Secondary analysis results	All sites using standard custom question; all sites using similar questions (varied by type of analysis)
Surveys		
Initial survey	• Site background • Site evaluation before ACSI • Reasons for joining the trans-NIH ACSI evaluation	All sites (57)
Final survey	• Intermediate outcomes • Longer term outcomes • Trans-NIH benefits	All sites (51)
Interviews		
Initial in-depth interview (primary focus: processes)	• Implementation process • Receipt and use of ACSI results • Trans-NIH benefits	Subset of sites (14 in 2005; 6 in 2006)
Final in-depth interview (primary focus: outcomes)	• Intermediate outcomes • Longer term outcomes	Subset of sites (20)
Final brief interview	• Benefits of ACSI use without full activities and data for full model	Subset of sites with less ACSI experience (5)
Observations		**Coverage (number of meetings)**
Observation of implementation and feedback meetings	How teams: • Implemented ACSI • Received and reacted to feedback	Sample of meeting types – implementation, initial feedback, follow-up feedback (15)
Observation of trans-NIH meetings	• Attendee questions and issues • Discussion topics • Case studies	All trans-NIH meetings (5)
Observation of leadership team meetings	• Management of trans-NIH effort • Perceptions about ACSI use across sites	Biweekly meetings (all meetings during evaluation period)

Results

The results are presented in relation to the four evaluation objectives:

- Objective 1: Through the offer of an ACSI license at no cost to participants, were Web teams encouraged to use an online customer satisfaction survey?
- Objective 2: What was the perceived value to the Web teams of using the ACSI?
- Objective 3: Did broad ACSI use provide additional enterprise-wide NIH benefits?
- Objective 4: Did the trans-NIH ACSI project provide any additional understanding about how NIH websites are used and are meeting NIH communication needs?

Web Team Participation Rates

Prior to the trans-ACSI project, only a handful of NIH Web teams were using online customer satisfaction surveys of any type. Three NIH organizations were using the ACSI survey method (for a total of seven websites). However a clear majority of NIH website representatives had indicated interest in using the ACSI, if funds permitted.

The central funding of the ACSI project allowed ACSI licenses to be offered to participating websites at no cost to them. The

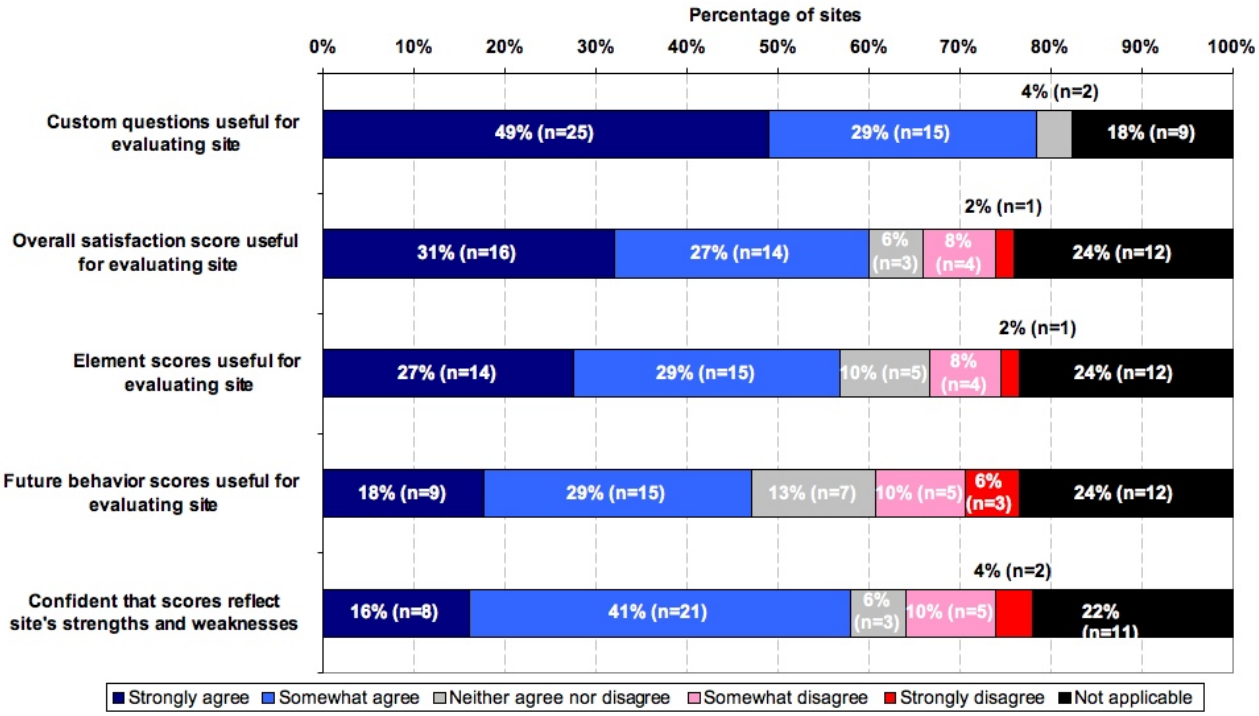
result was that all 60 of the website teams indicating preliminary interest signed up for the project. Of those original teams representing 60 websites, 55 sites actually implemented the ACSI, and 42 of those generated enough data to qualify for regular reporting of satisfaction scores (as of September 2006); 51 website teams completed the final survey; 5 of the original 60 withdrew for various reasons, such as inadequate Web traffic, changing priorities, or insufficient staff or management support. Low-traffic sites were the most likely to withdraw; these included Intranet sites and niche or specialty sites with very small target audiences or narrow topics.

The combination of the free ACSI license plus the significant support from the trans-NIH leadership team, the ACSI contractor, and the quarterly meetings were sufficient to increase NIH participation in the ACSI from seven websites to several times that.

Perceived Added Value of the ACSI

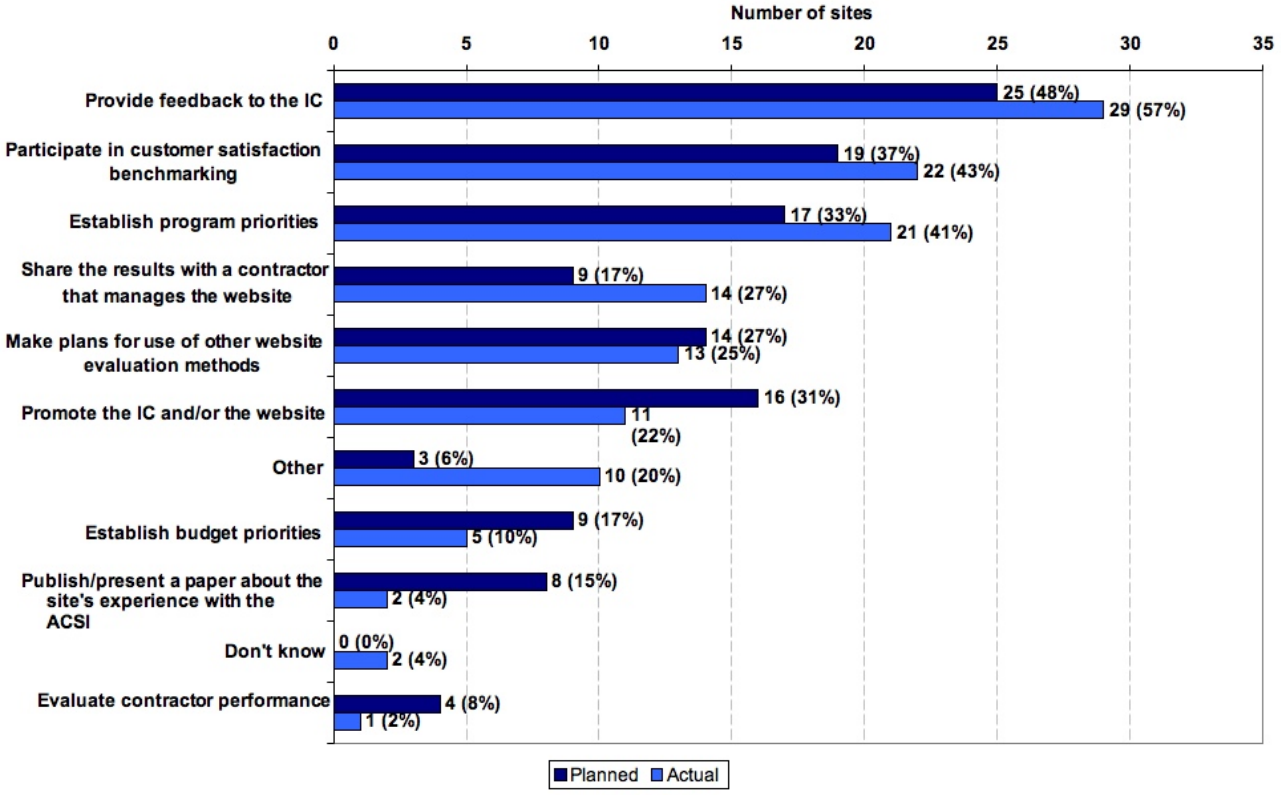
A major goal was to evaluate the use and value of the ACSI to NIH website teams. Based on the responses of 51 website teams, the respondents overwhelmingly (78%) strongly or somewhat agreed that the custom questions were useful for evaluating the website. About three fifths of respondents strongly or somewhat agreed that the overall customer satisfaction score and the element scores were useful. Respondents rated future behavior scores somewhat less useful, by comparison. A majority of respondents (57%) indicated confidence that scores reflected a website's strengths and weaknesses.

Figure 1. Usefulness of custom questions and ACSI survey scores as reported by participating NIH website teams (Method: Final Website staff survey, n=51)



The website teams were queried on their planned and actual use of the ACSI data. An overwhelming majority of respondents indicated that the ACSI data were more extensively used than planned to provide feedback to their NIH organization, to participate in customer satisfaction benchmarking, and/or to establish program priorities. Some responded that the ACSI data were shared with their website contractor, used to plan for use of additional evaluation methods, and/or used to promote the NIH organization or the website. For example, some NIH organizations used the positive results of their ACSI surveys to favorably promote their resources in annual reports [38], newsletters [39,40], congressional budget justifications [41], and reports to advisory groups [42,43]. A few used ACSI data to establish budget priorities, evaluate contractor performance, or publish or present a paper on the ACSI [44-49] (see also Multimedia Appendix 2 and Multimedia Appendix 4).

Figure 2. Use of ACSI survey data as reported by participating NIH website teams (Method: Initial Website staff survey, n=52, and final Website staff survey, n=51)



Website teams were asked what types of site improvements were planned based on what they learned from the ACSI data. The responses covered the breadth of possible website improvements. Almost half of the respondents cited site functionality and navigation. A third or more mentioned improved content, search, overall look and feel, and home page or subpage redesign. A handful mentioned site performance. For further details, see Multimedia Appendix 1 (pp. 3-20, Figure 3-13).

A clear majority (55%, 28/51 sites responding) indicated plans to use the ACSI data for their next website redesign; only a small minority (12%, 6/51) said they were not planning to use the ACSI data in the next redesign. However 25% were not sure (13/51); and one fifth said not applicable (7/51), which could imply that a site redesign was not anticipated.

Website teams were asked whether they were satisfied overall with the use of the ACSI to evaluate their website. The results indicate a roughly four to one balance of those agreeing versus disagreeing—67% (34/51) were strongly or somewhat satisfied, and 18% (9/51) were strongly or somewhat disagreed.

There is some indication that those website teams that actively used the ACSI data during the project were able to increase their overall ACSI customer satisfaction scores. For example, for the 12 websites that showed statistically significant changes in ACSI satisfaction scores, those sites that used the ACSI survey results for continuous website improvement and/or for evaluating effects of website changes tended to have higher satisfaction scores. Those sites that did not use the ACSI survey for those purposes tended to have lower satisfaction scores. These were the only conclusions that could credibly be drawn for the subset of websites with statistically significant changes in satisfaction scores. And those conclusions cannot be generalized to the entire group of participating websites given the absence of statistically significant data and the complexity of the survey and Web design processes.

The generally positive evaluative results need to be balanced by survey results that indicate significant constraints on the ability of Web teams to redesign their sites and to use and continue using the ACSI in the future. When asked about barriers to making changes to their website, almost half (47%, 24/51) of respondents mentioned staff time constraints, and about one quarter (27%, 14/51) noted financial resource constraints. About one fifth cited insufficient calendar time (16%, 8/51) or other reasons (12%, 6/51). Only 9 sites (18%) indicated that there were no barriers; 13 sites (25%) said that the question was not applicable, implying no plans to make major site changes.

Benefits of Trans-NIH ACSI Use

Another major goal was to evaluate the importance of the trans-NIH ACSI project to NIH as a whole. Based on interviews with a cross-section of Web teams (see Multimedia Appendix 1 for the interview guide) and observations of quarterly meeting discussions, the project greatly increased the focus on measurement of customer satisfaction with NIH websites. The project also encouraged a user-centered approach to NIH website design, improvement, and evaluation. In addition, the project strengthened the network of NIH website professionals and provided opportunities to share experiences and lessons learned and offer staff mentoring.

These results were a direct consequence of making the ACSI licenses available to all participating websites (basically, virtually all interested NIH websites), the nature of the ACSI process, which includes online reporting and periodic analytic support sessions (from ForeSee Results Inc.), and the quarterly trans-NIH meetings. Attendance at the quarterly meetings, held on the NIH main campus, ranged from about 30 to 60 persons and averaged about 45. The majority of websites had one or more team members present at most meetings. The demonstrated level of interest was usually high. Only 3 of 51 teams reporting had not sent a team representative to attend any quarterly meetings. Seven teams reported attendance at all meetings.

The NIH-wide meetings were especially helpful in highlighting contributions and challenges of the ACSI, contributing to an increased awareness and understanding of Web evaluation at NIH, and providing a forum to share lessons learned and identify future directions and opportunities. Web teams shared case studies of specific website experiences with the ACSI, including the use of different types of custom questions. For further details, see Multimedia Appendix 1 (pp. 4-5, Figure 4-1).

The trans-NIH project identified key factors associated with the successful use of the ACSI and with difficulties implementing the ACSI. Factors associated with success included the timing of the surveys with the website redesign cycle—the ACSI survey results were quite useful when planning a website redesign or in evaluating a completed website design. Also important is supportive management that believes in the value of customer surveys and Web evaluation in general. Another success factor is sufficient financial resources (in this project, for staff and website development costs—the cost of the ACSI survey itself was paid through central NIH funds).

Factors related to ACSI implementation difficulties included low-traffic websites. Based on the NIH experience, websites with fewer users, roughly anything less than 50,000 unique visitors per month, need to be monitored carefully to assure that enough completed survey responses are generated in a reasonable period of time. Low-traffic sites tended to include niche or specialty sites as well as Intranet sites, for which very high sampling rates may be needed, thus necessitating the use of persistent cookies to block repeat surveys for the same visitor (see below). Intranet websites with few or no outside users were likely to be problematic. For this NIH project, the Intranet sites had both low traffic and low survey response rates, which means it takes a long time to generate sufficient survey responses. Another factor associated with difficulties is a skeptical staff and/or management attitude toward surveys or Web evaluation in general. Infrequently, a technical issue, such as manual software coding to install the survey pop-up code, contributed to problems. This was the exception, however. The typical experience was easy technical implementation with automated software download and installation.

Another benefit of the trans-NIH approach was the approval of use of persistent cookies on NIH websites. Persistent means that the cookie was left on beyond the time of the initial site visit. The cookies did not collect any personally identifiable information and were used simply to block repeat surveys to the same visitor in a specified period of time (eg, 60 or 90 days). OMB policy generally prohibits use of cookies on federal government websites in order to assure that websites are not used to track individual Web use or collect personally identifiable information [50]. It is difficult to get an exception. But cookies can be used if there is a "compelling need," if privacy requirements are met, and if the cookie use has "personal approval by the head of the agency." NIH applied to the Secretary of Health and Human Services, who granted permissions because of the scope of the project and possible burden on the consumer (websites users) from repeat surveys. The cookies were used solely to block site visitors from receiving multiple surveys, and did not contain any personally identifiable information. The cookies helped alleviate concerns about visitors getting "survey weary" or, on the other hand, about a few visitors biasing the results by submitting multiple responses.

Additional Understanding about NIH Website Visitors

The use of a common survey method across a large number of NIH websites provided an opportunity to gain new insight or clarify earlier impressions about NIH Web visitors. The clear finding is that, overall, through its websites NIH serves multiple audiences with diverse information needs. Many NIH websites have significant percentages of health care provider, scientist, and consumer (including patients, families, and friends) visitors and provide information on a wide range of health, disease, treatment, research, and funding topics.

Based on responses to custom questions asked by 42 websites, students and patients each accounted for about one fifth of visitors, and health care professionals and scientists/researchers each accounted for about one seventh of visitors, on average. The general public (students, patients, families/friends, other) accounted for half to two thirds of visitors based on self-reported visitor roles. For further details, see Multimedia Appendix 1 (pp. 4-14, Figure 4-5).

Very few websites have earlier comparable survey data. For a handful of sites with earlier data, including MedlinePlus, TOXNET, Cancer.gov, and NHLBI, the results were reasonably consistent. The data from this trans-NIH study tended to confirm the trend over the last few years toward a large increase in consumer and general public use of NIH websites, in part due to greater emphasis by NIH on serving the general public's health information needs as well as needs of health care providers, scientists, and researchers.

Responses to custom questions asked by 31 websites indicated that, on average, the majority of visitors to NIH websites found the information they wanted. In response to the question "Did you find what you were looking for?" visitors responded: yes, 63%; no, 11%; still looking, 26%; partially, 21%; not sure, 9%; not looking for anything specific, 8%.

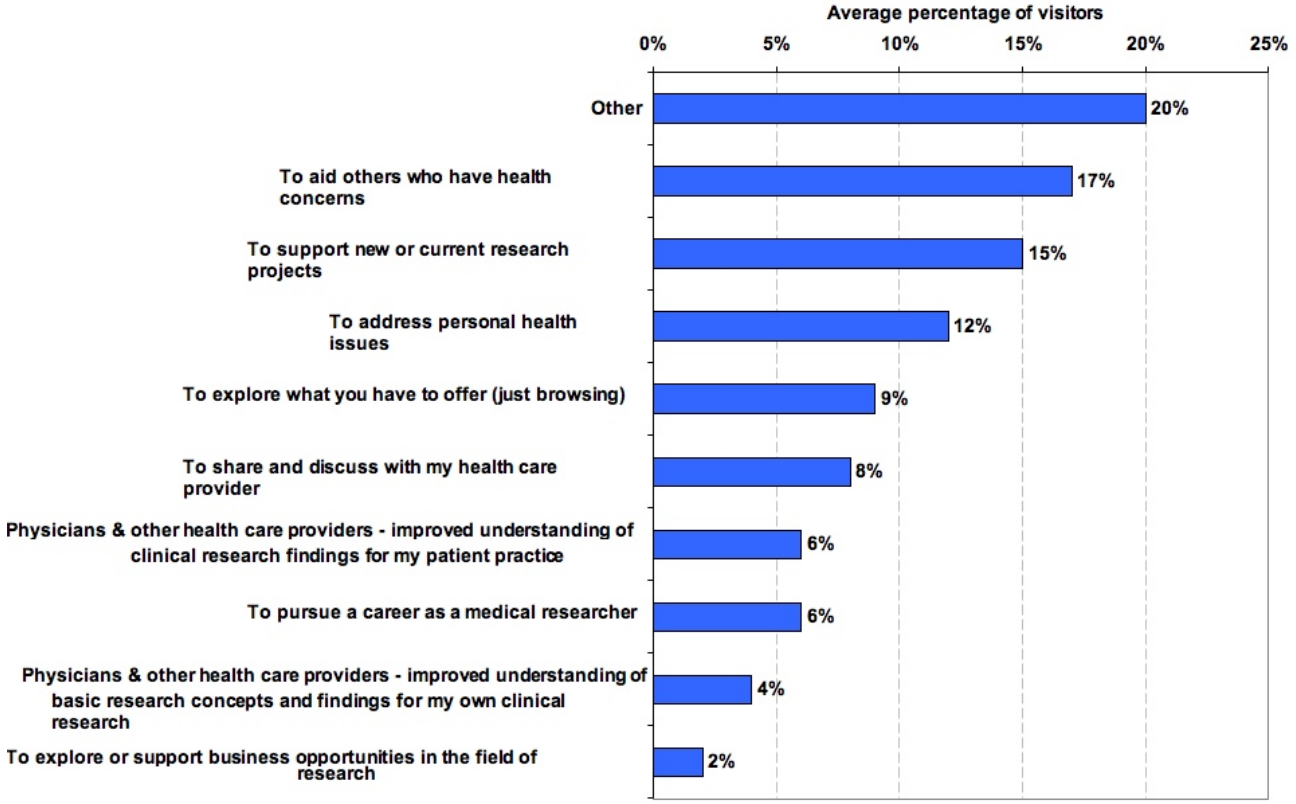
There were 26 sites using custom questions asking "How did you hear about (or get to) this site?" Across these sites, a search engine was cited most often (42%), followed by a link from another site (17%), and then by a bookmark (16%). For further details, see Multimedia Appendix 1 (pp. 4-16, Figure 4-7).

The trans-NIH leadership team did mandate one common custom question for all participating websites: "How do you

plan to use the information you find on this site today?" "You" in this context refers to the website user responding to the online survey. The ACSI contractor, ForeSee Results Inc., included this question on all custom surveys active in January 2006 (with the exception of sites that opted out); 35 sites included this trans-NIH question.

The results indicate a wide range of reported uses of information found on NIH websites. The response options selected indicate that while uses related to research and health practice are significant (about one quarter), there is an even greater emphasis on using information for personal health issues (about one third), whether for oneself or for family and friends. The one third combines the categories of aiding others who have health concerns, addressing personal health issues, and discussing personal health issues with a health care provider. This again reflects the shift in users since the advent of the Web, with a relatively large increase in patients and the public compared to the traditional (pre-Web) NIH core users from the research and health provider communities.

Figure 3. Intended use of information found on website as reported by site visitors (Note: Percentage of visitors indicating each intended use, averaged across all 35 reporting websites; percentages in this case add up to 100% because a standard question with the same response choices was used on all participating sites.) (Method: ForeSee Results Inc. standard custom question)



The use of the ACSI survey also provided a basis for benchmarking NIH websites against other federal government and private sector websites. The benchmarking is based on the combined responses to three ACSI standardized questions: "What is your overall satisfaction with this site?"; "How well does this site meet your expectations?"; "How does this site compare to your idea of an ideal Website?"

The customer satisfaction index can range from 0 to 100 based on a weighted average of responses to the three questions (which themselves use a 100-point Likert scale).

The NIH websites as a group scored consistently higher than the federal government and the private sector averages, based on 2006 quarter 4 data for US government websites [51] and 2006 annual data for private sector websites [52]. The average score of 81.3 for participating NIH websites compared very favorably with 73.9 for all federal e-government websites.

It should be noted that the NIH-wide customer satisfaction score varied during the course of the study depending on the number of sites participating and the relative performance of the sites included in the average. At the beginning and after the end of the study period, NIH scores were somewhat higher because some of the weaker performing websites had either not started up or had discontinued participation. The NIH-wide average quarterly score ranged from a high of 79 in 2004 quarter 4 and 81.3 in 2006 quarter 4 to a low of 75.1 in 2006 quarter 1, but in all quarters the NIH average was higher than the comparable federal e-government average score.

NIH average satisfaction scores also outpaced private sector scores. In the news/information sector in 2006 quarter 4, the average for all NIH was 81.6 compared to 72.9 for all e-government websites and 73.0 for all private sector websites using the ACSI. Leading individual websites in the news/information sector included the following, among NIH websites: MedlinePlus (NLM), 86; MedlinePlus in Spanish (NLM), 86; AIDSinfo (NLM), 84; NIDDK (National Institute of Diabetes and Digestive and Kidney Diseases), 84; and Cancer.gov in Spanish (NCI), 83. Among private sector

news/information websites, the leaders were as follows: USATODAY.com, 74; CNN.com, 73; ABCNEWS.com, 73; MSNBC.com, 72; and NYTimes.com, 72.

In the portal sector, in 2006 quarter 4, the NIH average satisfaction score was 80.8, the all e-government score, 74.9, and the private sector, 76.0. Leading individual NIH websites in the portal sector included Cancer.gov (NCI), 83; NHLBI (National Heart, Lung, and Blood Institute), 83; Office of Science Education/Office of NIH Director, 82; and NIAMS (National Institute of Arthritis and Musculoskeletal and Skin Diseases), 80. Leading private sector portal websites included Yahoo.com, 76; MSN.com (Microsoft Corp.), 74; and AOL.com (Time Warner Inc.), 74.

While the numeric customer satisfaction scores varied somewhat during the project, the NIH websites as a group scored consistently higher than e-government and private sector averages. The leading NIH websites individually scored significantly higher than the leading private sector websites in their class.

Aside from these global comparisons, it was not possible to conduct drill-down quantitative analyses of impacts on satisfaction scores. This was because, in the first instance, only 12 of the 55 websites implementing the ACSI showed statistically significant changes in satisfaction from start to finish. Second, while the ACSI standardized question responses give some indication of the most highly leveraged Web design changes, no quantitative data were collected on the specific Web changes made, if any, and their relationship to changes in satisfaction. Thus, while qualitative data based on interviews and surveys of Web teams are reported in this paper, drill-down quantitative analyses could not be credibly and validly carried out and, in any event, were beyond the project scope.

Discussion

This project was the first enterprise-wide ACSI application and probably the largest enterprise Web evaluation project to date in the US government. The project implemented the largest number of ACSI surveys (55) at any one government agency. Other agencies using the ACSI have multiple measures but in smaller numbers; for example, the Centers for Medicare and Medicaid Services are using 20, the US Department of State is using 15, the US Department of Agriculture uses 9, and the US Department of the Treasury uses 8 (personal communication, Ron Oberbillig, Federal Consulting Group, US Department of the Treasury, April 16, 2007).

The trans-NIH ACSI project met all of the original study and evaluation goals—a broad cross-section of NIH websites participated, the trans-NIH project leadership team drew from several NIH organizations and functioned very well for the 2-year project duration, NIH Web staff attendance at quarterly meetings was good to excellent, the project evaluation methodology was well designed and funded and fully implemented, and the evaluation itself was successful in identifying useful information on the site-specific and trans-NIH impacts of using the ACSI as well as assessing the success of the project as a whole.

Multimedia Appendix 5 is a PowerPoint presentation highlighting select evaluation and trans-NIH results, presented at the last trans-NIH meeting to be held as part of the project (October 2006). Multimedia Appendix 4 is a PowerPoint presentation discussing the enterprise-wide approach, presented at the Federal Consulting Group's ACSI Web Survey Group quarterly meeting (March 2007).

A majority of participating website teams reported significant benefits and new knowledge from the ACSI survey results and from being involved in the overall project process. The more experienced and better funded so-called "power users" among the participating NIH websites were able to use the ACSI as a ready-to-use customer satisfaction metric that provided pre-approved OMB clearance (a major advantage in streamlining the start-up process) and as a tool for incorporating custom questions into the survey in order to identify specific website issues and problems. Power users also employed the ACSI results as a source of information about site visitor demographics and as a means to analyze the satisfaction levels and information retrieval results of visitor subgroups to identify needed site improvements. The power users utilized the ACSI as a source of information for planning any follow-up or parallel work involving additional evaluation methods and as an archive of survey data for future use and analysis in website redesign and information enhancements.

These power users were able to apply the ACSI survey results to benchmark their particular NIH websites against other government and private sector websites and to gain insights about and opportunities for improving their Web presence through site-specific feedback. The ACSI results allowed power users to respond more quickly and effectively to the ever-evolving and changing Web environment and to help determine the impact of website changes and evaluate whether Web-based information dissemination programs are performing significantly better or worse over a defined period of time.

As a group, the participating NIH websites performed very well overall against US government and private sector benchmarks. The power user NIH websites—again, typically the larger and more heavily used, staffed, and funded websites—tended to have higher satisfaction scores than other participating websites. These websites also were more likely to use several evaluation methods in order to triangulate results and obtain more complete inferences and interpretations. However, with all NIH websites included, the NIH-wide average satisfaction score exceeded the government-wide average from the beginning of the project until the end.

As a consequence, NIH as a whole, and some individual NIH organizations, received significant positive media coverage of their Web performance during the course of the project [53-57]. Also, NIH received the first ever e-government award from the Federal Consulting Group / US Department of the Treasury—the Customer Performance Achievement Award—conferred by the OMB Administrator for Electronic Government and Information Technology in recognition of the success of the trans-NIH ACSI project.

Websites varied in their ability to implement the ACSI and utilize results. The majority of participating websites were able

to implement the ACSI and receive survey results, including satisfaction scores. Some sites were able to implement the ACSI but did not generate sufficient completed surveys to generate satisfaction scores due to low traffic on the website or because the ACSI was implemented too late in the study. However, these sites were able to obtain the results of their custom questions. The ACSI or any other online user survey does not work well with low-traffic websites. It simply takes too long to obtain a minimum sample for statistically significant results.

Due to the large number of websites involved, the trans-NIH project, out of necessity, implemented the ACSI in stages, determined in part by the degree of readiness of each website to participate. This generally meant that the more experienced better-staffed websites (including sites that had been pilot testing the ACSI) fully implemented the ACSI earlier and had more time to collect survey results. Other sites were not ready to implement the ACSI until late in the project. In addition, some sites that dropped out were replaced by others late in the project. The late starters in some cases did not have sufficient time to generate enough completed surveys.

Website teams that used the ACSI the longest tended to be satisfied with and find value in its use, especially for planning site changes and comparing versions of the website before and after revisions or redesigns. Teams with relatively later start dates and/or slow rates of collecting completed ACSI surveys were more likely to be dissatisfied with the ACSI because they did not have sufficient time or opportunity to receive and/or act on ACSI survey results.

Relative inexperience in using the survey may also have been related to perceived value because of the complexity of the survey results. The ACSI, unlike simpler survey methods, generates multidimensional results based on both standardized and custom questions. Segmentation of results, while analytically powerful, can also be daunting to the inexperienced.

In addition to time and experience, other key factors driving successful use of the ACSI or, by extension, other similar online survey methods, based on this project experience include staff and management buy-in, adequate resources, staff training and understanding, the website design cycle, and technical support.

Across all participating NIH websites, the Web teams derived substantially greater value from their custom question data and from segmentation data (breaking out results by specific types of visitors, information seeking goals, demographics, etc), than from the standardized ACSI questions. The custom question data provided many Web teams with valuable insight about visitor profiles and visit characteristics. For example, through cross-correlations between responses to custom and standardized questions, Web teams were able to identify visitor subgroups that were less satisfied and highlight needed website improvements. Many teams also took advantage of having a continuous source of customer feedback for tracking the visitor responses to website improvements implemented in response to ACSI data (as reflected in satisfaction scores).

The ACSI, like all online surveys in the Web environment, has relatively low response rates (typically about 5%, but ranging from 3% to 7%). The ACSI uses random intercepts and several cross-checks to help assure that nonresponse bias is minimized, but the latter is still a concern and warrants greater attention in the academic and survey research communities. NLM, NCI, and NHLBI, three of the participating NIH organizations, had used online surveys for several years prior to the ACSI. The prior surveys placed greater emphasis on the custom questions and less on standardized questions or benchmarking. Comparison of results about site visitors between the prior surveys and the ACSI results for several websites (eg, MedlinePlus, AIDSinfo, and TOXNET at NLM, and the NHLBI website) indicated that similar results were obtained between the earlier surveys and the ACSI surveys [22,23, personal communication, Cindy Love, April 30, 2007; personal communication, Mark Malamud, October 9, 2007]. This suggests that the ACSI survey results can be considered reasonably valid, and not unduly affected by non-response bias, unless there are undetected sources of non-response bias affecting all surveys over an extended time frame.

However, it is best not to rely too heavily on any one Web evaluation methodology. As noted earlier, a multidimensional approach is warranted and has been adopted by the more experienced better-funded NIH websites. The survey of NIH Web teams indicates that 21 of the participating teams practise, to varying degrees, a multidimensional approach. In addition to the ACSI, during the time of the trans-NIH project, 19 of the 21 websites also used Web log software, 18 used usability testing, 11 used expert or heuristic reviews, 4 used other types of surveys, 4 used focus groups, 3 used audience measurement and profiling, and 1 indicated other.

Conclusions

The trans-NIH leadership team believed in the importance of Web evaluation going into the trans-NIH ACSI project and was motivated to make the ACSI available to a broad group of NIH websites. The hope was to significantly increase the use of online customer surveys, the ACSI being a particular variant of the general class, within the NIH Web community. Further, the hope was that the project would not only increase NIH staff understanding of the value of this and other forms of Web evaluation, but also strengthen the management and financial support for Web evaluation at NIH.

The project was successful in increasing the use of and interest in online surveys and enhancing the understanding of the strengths and limitations of such surveys. A majority of participating websites found considerable added value in the survey process and results. However, many of the Web teams gave a clear indication in the project evaluation survey that notwithstanding the benefits, it was uncertain or questionable whether they would be able to fund the modest (US $20,000 or so per year per website) cost of renewing the ACSI from their own funds if central NIH funds were no longer available. As it turned out, central funding was not continued beyond the 2-year project life of this trans-NIH project, and each participating NIH website had to make its own decision whether to continue, and, if so, find its own funding to do so. The result was that only about one quarter of the NIH websites renewed their ACSI license, and half of those renewals were the early experimenters who had been using the ACSI for the longest time.

For this trans-NIH project, the US $18,000 survey license fee per website was considered to be competitive with other online survey options in terms of cost and to offer a better value added per dollar when considering the other benefits of the ACSI. For those websites wishing to continue, the FCG and ForeSee Results offered an ACSI "lite" version at US $15,000 (compared to US $25,000 for full service), but even at that price point there were relatively few renewals.

The NIH was fortunate to have the support of the Evaluation Set-Aside Program for the trans-NIH ACSI project. Much was learned, and many websites received significant added value, in their own estimation. But this was an experiment, not an ongoing operational activity. Without central funding, only the more experienced better-resourced larger websites, for the most part, continued with the ACSI.

Thus, a final lesson learned from the trans-NIH ACSI project experiment is the tenuous nature of Web evaluation in the age of e-government, when OMB and departmental policies are placing ever greater emphasis on Web-based delivery of government information and services. A parallel commitment to adequate evaluation of those Web-based activities may well be needed in order to help assure that the potential of the Web and other information technologies to improve customer and citizen satisfaction is fully realized.

Acknowledgments

The authors wish to thank the following persons who significantly contributed to the trans-NIH project reported in this paper: At ForeSee Results, Inc.: Larry Freed, President; Joel Van Haaften, Project Manager; Rick Jacobson, Director of Marketing; Errol Hau, Director of User Services; Bernie Lubran (formerly with FCG); and the several Customer Satisfaction Analysts who worked with NIH; At the Federal Consulting Group / US Department of the Treasury: Ron Oberbillig, Chief Operating Officer; and at NIH: the roughly 100 NIH staff that had responsibility for the NIH websites participating in the trans-NIH ACSI project.

This paper and the views expressed herein are, however, solely the responsibility of the authors.

This research was supported by funding from the National Institutes of Health (NIH) Office of the Director (OD), Evaluation Branch (EB), and by the individual NIH Institutes, Centers, and OD Offices that participated. The EB is now part of the Division of Evaluation and Systematic Assessment, NIH Office of Portfolio Analysis and Strategic Initiatives.

Authors' Contributions

All of the co-authors participated in biweekly trans-NIH leadership team meetings and were significantly involved in a wide range of issues and discussions on all aspects of the project. Co-authors Wood and Siegel had the lead responsibility for negotiating and managing the contracts with Westat Inc. and ForeSee Results Inc. and for coordinating with the Federal Consulting Group / US Department of the Treasury. Co-authors Wood, Siegel, Feldman, and Love provided substantial methodological and analytical guidance and content to the project. Co-author Rodrigues had the lead responsibility for communicating with the trans-NIH website staff via the NIH Web Authors Group and with the broader NIH community. Co-author Malamud had the lead responsibility for obtaining NIH and departmental approval of the use of a "persistent cookie" for the limited purpose of blocking repeat surveys to the same person. Co-author Lagana had the lead responsibility for working with the NIH Center for Information Technology on technical implementation and coordination issues. Co-author Crafts had the lead responsibility for planning and implementing the trans-NIH project evaluation contract.

Also, co-authors Feldman, Love, and Malamud represented NIH at the Federal ACSI Web Survey Users Group, and Love and Feldman serve on the Web Metrics Task Group of the Federal Web Manager's Advisory Council. Feldman, Love, and Malamud participated on the Leadership Team and had the practical experience of implementing and managing the ACSI for their NIH institutes both within and outside the scope of this project. Feldman managed the ACSI on five websites at NCI and helped coordinate an additional two; Love managed five at NLM, and Malamud managed two at NHLBI and helped coordinate an additional four.

This paper is based significantly on materials drafted by co-authors Wood, Siegel, Feldman, Love, and Crafts, and integrated and synthesized by Wood. These materials include presentations prepared by co-author Crafts for an October 4, 2006, trans-NIH ACSI meeting in Bethesda, MD; by co-authors Feldman and Love for a March 20, 2007, trans-Federal Government ACSI meeting in Washington, DC; and by co-authors Siegel and Wood for a January 22, 2007, symposium of the International Council for Scientific and Technical Information, London, England, UK. All three presentations are included as a Multimedia Appendix, along with the September 2006 Final Evaluation Report prepared by Dr. Crafts.

Co-author Love prepared the comprehensive references and bibliography and the ACSI methodology and historical sections.

All co-authors have reviewed and commented on earlier versions of this paper and on the draft evaluation report and variously on other source documents listed in the Multimedia Appendices.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Trans-National Institutes of Health (NIH) American Customer Satisfaction Index (ACSI) Web Site Evaluation. Final Report.

[PDF file (Adobe PDF), 1.2 MB - jmir_v10i1e4_app1.pdf]

Multimedia Appendix 2

Siegel ER, Wood FB. Customer satisfaction and performance metrics. Paper presented at: International Council for Scientific and Technical Information; January 22, 2007; London, UK.

[PPT file (Microsoft Powerpoint), 980 KB - jmir_v10i1e4_app2.ppt]

Multimedia Appendix 3

American Customer Satisfaction Index E-Government Satisfaction Index Quarterly Reports (Archived Links)

[PDF file (Adobe PDF), 64 KB - jmir_v10i1e4_app3.pdf]

Multimedia Appendix 4

Feldman S, Love CB. NIH's enterprise approach to measuring customer satisfaction. Paper presented at: ACSI User Group Meeting; March 20, 2007; Washington DC.

[PPT file (Microsoft Powerpoint), 1.2 MB - jmir_v10i1e4_app4.ppt]

Multimedia Appendix 5

Crafts J. Summary briefing on the goals, methods, and results of the Trans-NIH evaluation. Report presented at: Trans-NIH ACSI Meeting, October 4, 2006, Bethesda, MA.

[PPT file (Microsoft Powerpoint), 93 KB - jmir_v10i1e4_app5.ppt]

References

1. Lacroix EM, Mehnert R. The US National Library of Medicine in the 21st century: expanding collections, nontraditional formats, new audiences. Health Info Libr J 2002 Sep;19(3):126-132. [Medline: 12390234] [doi: 10.1046/j.1471-1842.2002.00382.x]
2. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005 Dec 12;165(22):2618-2624 [FREE Full text] [Medline: 16344419] [doi: 10.1001/archinte.165.22.2618]
3. Bright MA, Fleisher L, Thomsen C, Morra ME, Marcus A, Gehring W. Exploring e-Health usage and interest among cancer information service users: the need for personalized interactions and multiple channels remains. J Health Commun 2005;10 Suppl 1(7):35-52. [Medline: 16377599] [doi: 10.1080/10810730500265609]
4. Grama LM, Beckwith M, Bittinger W, Blais D, Lollar C, Middleswarth A, et al. The role of user input in shaping online information from the National Cancer Institute. J Med Internet Res 2005 Jul 1;7(3):e25 [FREE Full text] [Medline: 15998616] [doi: 10.2196/jmir.7.3.e25]
5. McCray AT, Dorfman E, Ripple A, Ide NC, Jha M, Katz DG, et al. Usability issues in developing a Web-based consumer health site. Proc AMIA Symp 2000:556-560. [Medline: 11079945]
6. Wood FB, Siegel ER, LaCroix EM, Lyon BL, Benson DA, Cid V, et al. A practical approach to e-government Web evaluation. IT Professional 2003 May;5(3). [doi: 10.1109/MITP.2003.1202231]
7. Sterne J. Web Metrics: Proven Methods for Measuring Website Success. New York: Wiley Publishing Inc; 2002.
8. Bader JL, Theofanos MF. Searching for cancer information on the internet: analyzing natural language search queries. J Med Internet Res 2003 Dec 11;5(4):e31 [FREE Full text] [Medline: 14713659] [doi: 10.2196/jmir.5.4.e31]
9. Redalen A, Miller N. Evaluating website modifications at the National Library of Medicine through search log analysis. D-Lib Magazine 2000 Jan;6(1). [doi: 10.1045/january2000-redalen]
10. Miller N. Analysis of user messages to MedlinePlus.gov. J Med Libr Assoc 2007 Jan;95(1):81-83 [FREE Full text] [Medline: 17252071]
11. Wood FB, Benson D, LaCroix EM, Siegel ER, Fariss S. Use of Internet audience measurement data to gauge market share for online health information services. J Med Internet Res 2005 Jul 1;7(3):e31 [FREE Full text] [Medline: 15998622] [doi: 10.2196/jmir.7.3.e31]
12. ; National Library of Medicine. MedlinePlus Quality Guidelines. 2006. NLM/NIH. URL: http://www.nlm.nih.gov/medlineplus/criteria.html [accessed 2007 Nov 7] [WebCite Cache ID 5TBBg4aOp]

13. Miller N, Lacroix EM, Backus JE. MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. Bull Med Libr Assoc 2000 Jan;88(1):11-17 [FREE Full text] [Medline: 10658959]
14. Wilson FL, Baker LM, Brown-Syed C, Gollop C. An analysis of the readability and cultural sensitivity of information on the National Cancer Institute's Web site: CancerNet. Oncol Nurs Forum 2000 Oct;27(9):1403-1409. [Medline: 11058972]
15. ; International Council on Active Aging. NIHSeniorHealth.gov: empowering older adults with health information. J Act Aging 2005 Mar;4(2):62-64 [FREE Full text] [WebCite Cache ID 5OCch6D4H]
16. Boyer C, Geissbuhler A. A decade devoted to improving online health information quality. Stud Health Technol Inform 2005;116:891-896. [Medline: 16160371]
17. ; Office of Management and Budget. Government Performance Results Act of 1993. whitehouse.gov. URL: http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.html [accessed 2007 Nov 7] [WebCite Cache ID 5OUlEYByh]
18. ; Office of Management and Budget. Circular No. A-130 Revised. Management of Federal Information Resources. whitehouse.gov. URL: http://www.whitehouse.gov/omb/circulars/a130/a130trans4.html [accessed 2007 Nov 7] [WebCite Cache ID 5OUjyxbYq]
19. ; Office of Management and Budget. Part IX: Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Notice; Republication. Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001: Public Law 106-554. Federal Register. 2002 Feb 22. (36) p. 8451-8460 URL: http://www.whitehouse.gov/omb/fedreg/reproducible2.pdf [accessed 2007 Nov 7] [WebCite Cache ID 5OUl1Zm2n]
20. ; Web Managers Advisory Council, Interagency Committee on Government Information (2004). Recommended Policies and Guidelines for Federal Public Websites. Final Report of the Interagency Committee on Government Information. Submitted to The Office of Management and Budget June 9, 2004. usa.gov. URL: http://www.usa.gov/webcontent/about/documents/icgi_report.html [accessed 2007 Nov 19] [WebCite Cache ID 5O6yJ5H4c]
21. ; Memorandum for the Heads of Executive Departments and Agencies. December 17, 2004. Policies for Federal Agency Public Websites. Washington, DC: Office of Management and Budget, Executive Office of the President; 2004. URL: http://www.whitehouse.gov/omb/memoranda/fy2005/m05-04.pdf [accessed 2007 Nov 19] [WebCite Cache ID 5O6wBWch4]
22. ; National Library of Medicine's MedlinePlus Survey Results 2001. NLM/NIH. URL: http://www.nlm.nih.gov/medlineplus/survey/index.html [accessed 2007 Nov 19] [WebCite Cache ID 5TTQuFQbV]
23. ; MedlinePlus Survey Results 2003. NLM/NIH. URL: http://www.nlm.nih.gov/medlineplus/survey2003/index.html [accessed 2007 Nov 19] [WebCite Cache ID 5TTQwmINj]
24. Fornell CG, Johnson MD, Anderson EW, Cha J, Bryant BE. The American Customer Satisfaction Index: nature, purpose and findings. J Mark 1996;60(4):7-18. [doi: 10.2307/1251898]
25. Stowers GNL. Measuring the Performance of E-Government. Washington DC: IBM Center for the Business of Government; 2004. URL: http://www.businessofgovernment.org/pdfs/8493_Stowers_Report.pdf [accessed 2007 Nov 19] [WebCite Cache ID 5ODuZ16VL]
26. Freed L. American Customer Satisfaction Index E-Government Satisfaction Index. Ann Arbor, MI: ForeSee Results; 2003. URL: http://www.fcg.gov/pdf/e-gov_comm_092003.pdf [accessed 2007 Nov 19] [WebCite Cache ID 5ODsndOtY]
27. Yao ML. The President's Management Council: An Important Management Innovation. Arlington, VA: PricewaterhouseCoopers Endowment for The Business of Government; 2000. URL: http://www.businessofgovernment.org/pdfs/yao.pdf [accessed 2007 Nov 19] [WebCite Cache ID 5TTRN1kIb]
28. ; US General Services Administration. Vice President Launches First Governmentwide Customer Satisfaction Survey. 1999 May 26. gsa.gov. URL: http://www.gsa.gov/Portal/gsa/ep/contentView.do?contentType=GSA_BASIC&contentId=9177&noc=T [accessed 2008 Feb 6]
29. Fornell C. ACSI Commentary: Federal Government Scores. 2004 Dec 14. theacsi.org. URL: http://www.theacsi.org/index.php?option=com_content&task=view&id=152 [accessed 2007 Nov 19] [WebCite Cache ID 5VQkS2ygX]
30. Fornell C. ACSI Commentary: Federal Government Scores. 2005 Dec 15. theacsi.org. URL: http://www.theacsi.org/index.php?option=com_content&task=view&id=151&Itemid=62 [accessed 2007 Nov 19] [WebCite Cache ID 5O6qiDnqJ]
31. Fornell C. ACSI Commentary: Federal Government Scores. 2006 Dec 15. theacsi.org. URL: http://www.theacsi.org/index.php?option=com_content&task=view&id=162&Itemid=62 [accessed 2007 Nov 19] [WebCite Cache ID 5O6qmWYOo]
32. ; Office of Portfolio Analysis and Strategic Initiatives. Evaluation Branch. OPASI/NIH. URL: http://opasi.nih.gov/desa/eb/ [accessed 2008 Feb 3] [WebCite Cache ID 5ODybC1og]
33. ; Office of Portfolio Analysis and Strategic Initiatives. Evaluation Branch. NIH Evaluation Set-Aside Program. OPASI/NIH. URL: http://opasi.nih.gov/desa/eb/setaside.asp [accessed 2008 Feb 3] [WebCite Cache ID 5ODxxciEa]
34. ; ForeSee Results home page. ForeSee. URL: http://www.foreseeresults.com/ [accessed 2007 Nov 19] [WebCite Cache ID 5OE0vWASo]
35. ; US Department of the Treasury. Federal Consulting Group home page. fcg.gov. URL: http://www.fcg.gov/ [accessed 2007 Nov 19] [WebCite Cache ID 5OE16iVsm]
36. Fornell C, Van Amburg D, Morgeson F, et al. The American Customer Satisfaction Index at Ten Years, ACSI 1994-2004. A Summary of Findings: Implications for the Economy, Stock Returns and Management. Ann Arbor, MI: University of Michigan, Stephen M. Ross School of Business, National Quality Research Center; 2005.

37. ; American Customer Satisfaction Index Methodology Report. Ann Arbor, MI: University of Michigan, Stephen M. Ross School of Business, National Quality Research Center; 2005.
38. ; National Institutes of Health, National Library of Medicine. Programs and Services FY 2005. Bethesda, MD: US Department of Health & Human Services; 2005. URL: http://www.nlm.nih.gov/ocpl/anreports/fy2005.pdf [accessed 2007 Nov 19] [WebCite Cache ID 5OUoYTJN4]
39. ; National Cancer Institute. NCI Web site rated best e-government main site. NCI Cancer Bull. 2005 Jan 4. (1) URL: http://www.cancer.gov/ncicancerbulletin/NCI_Cancer_Bulletin_10405/page9 [accessed 2007 Nov 19] [WebCite Cache ID 5O9YbHIHw]
40. ; National Institute of Diabetes and Digestive and Kidney Diseases (2005). NIDDK health information Web sites rate high in customer satisfaction. In: NIDDK Director's Update (June 15, 2005). NIDDK/NIH. URL: http://www2.niddk.nih.gov/AboutNIDDK/Director/Update/Archives/20050615.htm [accessed 2007 Nov 19] [WebCite Cache ID 5TTVngmhZ]
41. ; National Library of Medicine. FY 2006 Budget. NLM/NIH. URL: http://www.nlm.nih.gov/about/2006CJ.pdf [accessed 2008 Feb 6] [WebCite Cache ID 5OUpLHp6T]
42. Wood F, Lacroix EM. NLM User Studies. In: Minutes of the Board of Regents February 10-11, 2004. Bethesda, MD: National Library of Medicine; 2004. p. 9-11 URL: http://www.nlm.nih.gov/od/bor/bor204.pdf [accessed 2007 Nov 19] [WebCite Cache ID 5OUpeqLEj]
43. ; National Institute of Dental and Craniofacial Research. Satisfaction with NIDCR Web site compares favorably with other government and private sector sites. In: Director's Report to the National Advisory Dental and Craniofacial Research Council. June 2005. NODCR/NIH. URL: http://www.nidcr.nih.gov/NewsAndReports/ReportsPresentation/DirectorsReportJune2005.htm [accessed 2007 Nov 19] [WebCite Cache ID 5TTWcAkMI]
44. Fariss S, Love CB. Now you've built it…Does it work? Paper presented at: E-Government Citizen Satisfaction Summit; September 21, Washington, DC; 2004 [FREE Full text] [WebCite Cache ID 5OCPkAcWn]
45. Feldman S. Evaluation of NCI Website. Paper presented at: E-Government Citizen Satisfaction Summit; September 21,. Washington, DC; 2004. URL: http://www.fcg.gov/documents/e-gov_satisfaction_summit.ppt [accessed 2007 Nov 19] [WebCite Cache ID 5OCPyaVRs]
46. Martin KL, Feldman S. Who's behind the screen & what are they looking for? Paper presented at: National Cancer Institute HINTS Data Users Conference; January 20, 2005, St. Pete Beach, FL. URL: http://hints.cancer.gov/hints/docs/Martin_Feldman.pdf [accessed 2007 Nov 19] [WebCite Cache ID 5O9bFGzID]
47. Feldman S. Demonstrating the value, impact and return on investment from evaluating Web-based services: An overview of strategies. In: Paper presented at: Digital Government Institute. Website Evaluation: Making Web Services Integral to Agency Mission; November 7-8, 2005; Washington, DC.
48. Rodrigues D. CSI survey results and how they are used. In: Paper presented at: Healthfinder Steering Committee Meeting; December 11, 2006; Rockville, MD.
49. Cadden C. How physicians are using the federal Treatment Guidelines: Data from AIDSinfo's online American Customer Satisfaction Index survey. In: Paper presented at: American Conference for the Treatment of HIV (ACTHIV); May 31 to June 3 2007; Dallas, TX.
50. ; Office of Management and Budget. Memorandum for the Heads of Executive Departments and Agencies. June 22, 2000. Privacy Policies and Data Collection on Federal Web sites. whitehouse.gov. URL: http://www.whitehouse.gov/omb/memoranda/m00-13.html [accessed 2007 Nov 19] [WebCite Cache ID 5OFiCp4VE]
51. Freed L. American Customer Satisfaction Index E-Government Satisfaction Index. Ann Arbor, MI: ForeSee Results: 2006 Dec 15. URL: http://www.fcg.gov/documents/acsi-results-12-2006.pdf [accessed 2007 Nov 19] [WebCite Cache ID 5O6kzzepa]
52. Freed L. American Customer Satisfaction Annual E-Business Report. Ann Arbor MI: ForeSee Results; 2006. URL: http://www.foreseeresults.com/_downloads/acsicommentary/ACSI_E-Business_Report_Aug06.pdf [WebCite Cache ID 5OUqj4X9n]
53. Chabrow E. High marks for government jobs and health sites: the University of Michigan's quarterly customer-satisfaction survey gives some government sites higher scores than commercial sites. InformationWeek. 2004 Mar 15. URL: http://www.informationweek.com/story/showArticle.jhtml?articleID=18312139 [accessed Nov 19] [WebCite Cache ID 5OFphz0yg]
54. Michael S. Consumers wowed by Web sites. FCW.com. 2004 Mar 21. URL: http://www.fcw.com/print/10_7/news/82380-1.html [accessed 2007 Nov 19] [WebCite Cache ID 5OFpXiGwT]
55. Pulliam D. Customer satisfaction with federal Web sites nudges up a notch. GovernmentExecutive.com. 2005 Jun 13. URL: http://www.govexec.com/dailyfed/0605/061305p1.htm [accessed 2007 Nov 19] [WebCite Cache ID 5OFqRBAUz]
56. Perera D. Health sites spur satisfaction. Government Health IT. 2005 Jun 14. URL: http://govhealthit.com/article89241-06-14-05-Web [WebCite Cache ID 5OFrUIB8P]
57. ; NIH Web sites win customer satisfaction. 2007 Jan 12. (1) p. 12 URL: http://nihrecord.od.nih.gov/pdfs/2007/01122007_Record.pdf [accessed 2007 Nov 19] [WebCite Cache ID 5O9bnPL7v]

Abbreviations

ACSI: American Customer Satisfaction Index
GSA: Government Services Administration
NCI: National Cancer Institute
NHLBI: National Heart Lung and Blood Institute
NIH: National Institutes of Health
NLM: National Library of Medicine
OMB: Office of Management and Budget

submitted 05.06.07; peer-reviewed by B Bailey; comments to author 19.06.07; revised version received 09.10.07; accepted 23.10.07; published 15.02.08.

Please cite as:
Wood FB, Siegel ER, Feldman S, Love CB, Rodrigues D, Malamud M, Lagana M, Crafts J
Web Evaluation at the US National Institutes of Health: Use of the American Customer Satisfaction Index Online Customer Survey
J Med Internet Res 2008;10(1):e4
URL: http://www.jmir.org/2008/1/e4/
doi:10.2196/jmir.944
PMID:18276580

Original Paper

Reliability of Internet- Versus Telephone-Administered Questionnaires in a Diverse Sample of Smokers

Amanda L Graham[1], PhD; George D Papandonatos[2], PhD

[1]Department of Oncology, Georgetown University Medical Center / Cancer Control Program, Lombardi Comprehensive Cancer Center, Washington, DC, USA

[2]Center for Statistical Sciences, Brown University, Providence, RI, USA

Corresponding Author:
Amanda L Graham, PhD
Georgetown University Medical Center / Lombardi Comprehensive Cancer Center
3300 Whitehaven St, NW
Milton Harris Bldg, Suite 4100
Washington, DC 20007
USA
Phone: +1 202 687 7905
Fax: +1 202 687 0305
Email: alg45@georgetown.edu

Abstract

Background: Smoking is more prevalent among lower-income individuals and certain racial/ethnic minorities. Addressing tobacco cessation among diverse populations is an urgent public health priority. As Internet use continues to rise among all segments of the US population, Web-based interventions have enormous potential to reach priority populations. Conducting Web-based smoking cessation research in priority populations requires psychometrically sound measurement instruments. To date, only one published study has examined the psychometric properties of Internet-administered measures commonly used in Web-based cessation trials. However, the sample was homogeneous with regard to race/ethnicity and income. We sought to replicate and extend these findings in a more diverse sample of smokers.

Objective: The aim was to examine the internal consistency and test-retest reliability of measures commonly used in smoking cessation clinical trials among racial/ethnic minorities and smokers with lower income.

Methods: Participants were enrolled in a randomized trial of the efficacy of an Internet smoking cessation program between June 2005 and September 2006. Following a baseline telephone assessment and randomization into the parent trial, participants were recruited to the reliability substudy. In phase I of recruitment, all participants in the parent trial were recruited to the substudy; in phase II, all consecutive racial/ethnic minority participants in the parent trial were recruited. Race and ethnicity were assessed via self-report using two standard items from the US Office of Management and Budget. An email was sent 2 days after the telephone assessment with a link to the Internet survey. Measures examined were quit methods, perceived stress, depression, social support, smoking temptations, alcohol use, perceived health status, and income. Internal consistency and test-retest reliability of Internet- versus telephone-administered measures were examined within four strata defined by race/ethnicity (non-Hispanic White, racial/ethnic minority) and annual household income (US $40,000 or less, more than $40,000).

Results: Of the 442 individuals invited, 319 participated (72% response rate): 52.4% were non-Hispanic White, 22.9% Black, 11.6% Hispanic, 7.8% Asian, 4.4% American Indian / Alaska Native, and 1% Native Hawaiian / Other Pacific Islander. About half (49.4%) reported an annual household income of US $40,000 or less, and 25.7% had a high school degree or less. Test-retest reliability was satisfactory to excellent across all strata for the majority of measures examined: 9 of 12 continuous variables had intraclass correlation coefficients ≥ 0.70, and 10 of 18 binary variables and both ordinal variables had kappa coefficients ≥ 0.70. Test-retest reliability of several quit methods varied across strata.

Conclusions: Race/ethnicity and income do not affect the psychometric properties of most Internet-administered measures examined. This knowledge adds to the confidence of conducting Web-based smoking cessation research and strengthens the scientific rigor of collecting information via the Internet on racial/ethnic minority and low-income subgroups.

Trial registration: clinicaltrials.gov NCT00282009 (parent trial)

(J Med Internet Res 2008;10(1):e8) doi:10.2196/jmir.987

KEYWORDS

Reliability; smoking; Internet; diversity; measurement; psychometrics; minority groups; questionnaires; socioeconomic factors; social class; poverty; African Americans; Hispanic Americans

Introduction

Although the overall prevalence of smoking has declined in recent years, now at 20.9% among US adults, smoking continues to be more prevalent among individuals with lower levels of income and education and among certain subgroups of racial/ethnic minorities [1]. For instance, smoking prevalence is 29.9% among those living below the poverty line, 43.2% among adults with a General Educational Development (GED) diploma, 32.6% among those with 9-11 years of education, 26.7% among African American men, and 32% among American Indians / Alaska Natives [1]. In addition, low education and income also have been linked to lower rates of quit attempts and quit success [2,3]. Given the enormous health burden and economic impact of smoking [4,5], addressing tobacco cessation among diverse populations has been identified as an urgent public health priority [6].

Increasingly, the Internet is being recognized as having great potential to address disparities in health and health risk behaviors (such as tobacco use) by providing information, treatment, and support to traditionally underserved populations [7-12]. More than 70% of US adults now use the Internet [13], and online usage has increased steadily since 2000 across race, education, income, age, and rural/urban categories [14,15]. In 2005, a majority of African Americans (57%) and Latinos (70%) reported using the Internet, as did 49% of individuals living in households with an annual income of less than US $30,000 [14]. In addition to the reach of the Internet, its 24/7 availability, the ability to engage with others as anonymously as desired, and the use of audio, video, and numerous other interactive features make it an appealing dissemination channel for health information and behavior change interventions. Indeed, with thousands of health-related websites in existence, the Internet now plays a meaningful role in the health care system, often serving as the primary source of health-related information and support for consumers.

The use of the Internet among smokers has increased steadily in recent years as well. In 2006, 9% of online adults (more than 10 million people) had searched the Internet for help in quitting smoking [16], up from 6% in 2002 [17]. Studies of Web-based cessation programs are growing rapidly in number [18], with early studies describing the development, usability, and pilot testing of programs [19-24] and more recent reports describing randomized efficacy trials [25-32]. To date, the majority of these studies have focused on “mainstream” Internet users who are largely non-Hispanic White, college educated, and have higher incomes. However, given the growth of Internet use across all demographic subgroups and the recent national attention on eliminating health disparities [33-37], research and development efforts will need to increasingly focus on tobacco use among priority populations such as racial/ethnic minorities [38] and those with lower levels of income and education.

Despite the overall increase in Internet use, it must be acknowledged that access to Web-based cessation programs is still uneven across populations with regard to income and race/ethnicity, with the poor and racial/ethnic minorities having more limited access. However, the persistence of a “digital divide” does not negate the need to conduct rigorous efficacy and effectiveness studies in these subgroups. Rather, it underscores the importance of research to understand for whom, why, and under what conditions Internet cessation programs are effective and to elucidate new directions to further reduce the digital divide.

Critical to the conduct of Web-based cessation research with more diverse populations will be the availability of measurement instruments that have been validated using samples of the target audience [39-41]. The assumption of universal applicability of standardized scales normed on majority populations needs to be explicitly tested across domains (such as racial/ethnic background, income, and education) to ensure that their use with specific subgroups is relevant and appropriate [41]. A growing body of evidence suggests that the reliability and validity of data obtained using questionnaires administered via the Internet are generally consistent with results obtained through paper-and-pencil and computer-administered questionnaires. However, the majority of these studies employ between-group comparisons. Cross-method consistencies examined within subjects have been demonstrated for several constructs, including dietary intake [42], independent life skills among youth [43], health status and health behaviors [44], and psychopathology screening [45].

To date, only one published study that we know of has examined the cross-method consistency of Internet- and telephone-administered measures commonly used in smoking cessation clinical trials using a within-subject design [46]. Our research group found that the internal consistency and test-retest reliability coefficients were comparable for Internet- and telephone-administered measures of stress, depression, self-efficacy, social support, perceived health status, alcohol use, and previous quit methods [46]. However, the sample in this ongoing study was primarily non-Hispanic White (80%) with a household income above US $30,000 (73%). It is important not only to determine that assessment instruments perform adequately when administered via the Internet, but also that they demonstrate sound psychometric properties across subgroups when administered via the Internet [47,48]. Therefore, the goal of the present study was to replicate and extend these findings in a more diverse sample of smokers. Specifically, we were interested in determining whether the psychometric properties of the measures previously examined were comparable across categories of race/ethnicity and income when administered online and by telephone.

Methods

Sample Recruitment

Participants were enrolled in a parent study that is an ongoing randomized controlled trial of the efficacy of an Internet smoking cessation program (QuitNet) and telephone counseling (Clinicaltrials.gov: NCT00282009). Recruitment into the parent trial has been described elsewhere [28]. Following a baseline telephone assessment, participants were randomized to treatment and invited to participate in a reliability substudy. Those who agreed were emailed 2 days later with a link to the online survey. Each participant's unique study identification number was embedded into the link to the online survey so that responses could be joined with their telephone survey data. A description of the online survey administration is available in Graham et al [46]. Participants were paid US $15 for completing the online survey.

Recruitment to the substudy was conducted in two phases. In phase I (June to September 2005), all individuals randomized to the parent trial were recruited. This yielded a sample that was primarily non-Hispanic White with a household income above US $30,000. To increase the heterogeneity of the sample, all racial/ethnic minority participants consecutively randomized to the parent trail were recruited in phase II (October 2005 to September 2006). Race and ethnicity were assessed using the US Office of Management and Budget [49] 2-question format. Participants were first asked to indicate their race from one of five categories: (1) American Indian or Alaska Native: a person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment; (2) Asian: a person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam; (3) Black or African American: a person having origins in any of the black racial groups of Africa; (4) Native Hawaiian or Other Pacific Islander: a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands; and (5) White: a person having origins in any of the original peoples of Europe, the Middle East, or North Africa. Next, participants were asked to indicate if they were Hispanic or Latino, meaning a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. These categories are required in all federally funded research studies in the United States. The study received human subject protections approval from the Georgetown University Medical Center institutional review board.

Measures

In the parent trial, the baseline telephone assessment included measures of demographic, smoking, and psychosocial characteristics. To be sensitive to response burden on participants in an Internet-based trial, brief measures and items from large national epidemiologic surveys with known psychometric properties were selected. The present study examined the reliability of the following subset of measures administered via the Internet.

Smoking Temptations Questionnaire (Short-Form)

The short-form (9-item) version of the Smoking Temptations Questionnaire [50] assessed the temptation to smoke in different situations. Each item is rated on a 5-point scale ranging from 1 "not at all tempting" to 5 "extremely tempting." The questionnaire can be scored to form a total score, as well as three subscale scores that measure temptations in positive affect or social situations, negative affect situations, and habitual or craving situations. This short form is derived from a 17-item measure for which internal consistency coefficients are as follows: Positive Affect / Social (6 items, Cronbach alpha = 0.857), Negative Affect (6 items, Cronbach alpha = 0.946), and Habit/Addictive (5 items, Cronbach alpha = 0.800) [50].

Partner Interaction Questionnaire

Supportive behaviors from a spouse/partner have been shown to predict successful quitting [51,52], and negative behaviors predict relapse [53,54]. The Partner Interaction Questionnaire (PIQ) [53] is the most commonly used measure of spouse/partner support related to cessation. We administered a modified version of the PIQ that measures the receipt of specific behaviors from the person who follows the participant's efforts to quit smoking most closely, not just a spouse/partner [55,56]. The modified version assessed how frequently the participant's support person exhibited three positive and three negative behaviors [46], with responses of never (0), almost never (1), sometimes (2), fairly often (3), and very often (4). The three positive items were "express pleasure at your efforts to quit," "congratulate you for your decision to quit smoking," and "express confidence in your ability to quit/remain quit." The three negative items were "mention being bothered by smoke," "ask you to quit smoking," and "criticize your smoking." Cronbach alpha coefficients were 0.92 for the 3-item positive subscale and 0.84 for the 3-item negative subscale.

Perceived Stress Scale

Stress has been implicated in problems quitting smoking and in relapse [57]. The 4-item Perceived Stress Scale (PSS) [58] assessed the degree to which participants found their lives to be unpredictable and uncontrollable during the past month. Response options were never (0), almost never (1), sometimes (2), fairly often (3), and very often (4). Cronbach alpha reliability coefficients range from 0.60 to 0.72 [58,59]. Test-retest correlations range from 0.85 over 2 days in a college sample to 0.55 over 6 weeks in a smoking cessation sample [58].

Center for Epidemiological Studies Depression Scale

Symptoms of current depression were measured using the 10-item Center for Epidemiological Studies Depression Scale (CES-D) [60]. Scores on the CES-D have been positively associated with smoking prevalence and intensity and failure to quit in representative samples of US adults [61]. The CES-D is widely used in smoking cessation trials in the United States and abroad (eg, [62-67]). Each item is rated on a 4-point scale to indicate the frequency of occurrence during the past week. Response options were modified to less than one day (0), one to two days (1), three to four days (2), and five to seven days (3). Test-retest correlations range from 0.21 to 0.84, with an

overall correlation of 0.71, at an average time interval of 22 days [60].

Alcohol Use

Alcohol use is a common barrier to cessation [68,69]. Participants were first asked if they drank any alcohol. Using items from the then current 2002 Behavioral Risk Factor Surveillance System [70], those who said yes were asked to indicate how many days per week on average they drank alcohol, how many drinks they typically had on a drinking day, and the maximum number of drinks they had on one occasion during the past month. In addition, we used a slightly modified version of a 2-item screener [71] to assess problems associated with alcohol use. The original questions asked about alcohol and drug use conjointly; our modification dropped the wording about other drugs so that questions read as follows: "In the last year, have you had more to drink than you meant to?" and "In the last year, have you felt you wanted or needed to cut down on your drinking?" These items have high specificity (80%-90%) to detect current alcohol problems.

Quit Methods

Participants indicated whether they had ever used various methods to quit smoking, including cold turkey, pamphlet or book, individual counseling, group counseling, nicotine patch, nicotine gum, nicotine nasal spray, nicotine lozenge, nicotine inhaler, Zyban (bupropion), switching to chewing tobacco or snuff, an Internet program (not including QuitNet), telephone counseling, acupuncture, hypnosis, or any other method.

Perceived Health Status and Medical History

Using the item from the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36), participants rated their current health status on a 5-point scale from 1 (excellent) to 5 (poor) [72]. Participants were also asked if they had ever had a smoking-related illness (yes/no).

Income

Income is considered a sensitive question that some participants may not be comfortable answering. We examined its reliability to determine if the greater anonymity of the Internet would result in different responses than telephone administration. Total household income during the past year was assessed with eight response options: less than US $10,000, $10,000-20,000, $20,000-30,000, $30,000-40,000, $40,000-50,000, $50,000-75,000, $75,000-100,000, and $100,000 or more.

Statistical Analysis

The first set of analyses documents the recruitment process and describes the recruited sample, including a comparison of participation rates between the original study and the present study. To examine the generalizability of the final sample, we characterized survey participants on a range of demographic, smoking, and psychosocial variables. Frequency tables are used to summarize the categorical data, and both parametric and nonparametric tests are employed to determine the statistical significance levels.

The test-retest reliability of measures across modes of survey administration (Internet versus telephone) was examined by race/ethnicity and income. Specifically, we conducted stratified analyses that compared and contrasted (1) non-Hispanic White participants versus racial/ethnic minorities and (2) low-income versus high-income participants. The group of racial/ethnic minorities is comprised of participants who reported their race as African American, Asian, Native Hawaiian / Other Pacific Islander, or American Indian / Alaska Native, or their ethnicity as Hispanic. Based on a naturally occurring median split, the binary income variable was created with low income representing US $40,000 or less (49.2%) and high income representing more than US $40,000 (50.8%). We considered using educational level as a stratification variable instead of income but decided against it. Since only 21% of our subjects had a high school degree or less, the uneven sample size would have resulted in low power for testing differences between Internet- and telephone-administered measures among subjects with lower educational level, as well as imprecise estimates of the corresponding reliability coefficients.

In Table 1, the test-retest reliability of all continuous variables is examined across survey methods using the intraclass correlation coefficient (ICC), calculated according to formula ICC(3,1) of Shrout and Fleiss [73]. This version of the ICC measures the correlation between a single rating on a continuous measure using the Internet survey, with a single rating of the same measure obtained over the telephone, when Internet and telephone are the only channels of interest for administering the survey (fixed rater scenario). In large samples, the ICC has an F distribution that can be used to derive asymptotic 95% confidence interval (95% CI) estimates. Test-retest reliability above 80% is usually sought in method comparisons, with 70% considered an acceptable value.

Since reliability measures are based upon mean-centered versions of the variables of interest, they are insensitive to participants' tendencies to provide consistently higher responses on one survey instrument than another. Therefore, examination of test-retest reliability for these two survey methods was supplemented by *t* tests aimed at detecting the presence of any systematic bias as manifested by location differences between the Internet and telephone surveys. We report the results of these *t* tests below, but these data have been omitted from Table 1 due to space limitations (the complete set of tables is available upon request from the corresponding author). To allow for the presence of outliers in the data, robust location tests based on the Wilcoxon statistic were also carried out. Additionally, effect size measures based on standardized mean differences were estimated for each stratum, allowing us to distinguish clinically significant from merely statistically significant results. With approximately 160 subjects per stratum of interest, this study was designed to ensure detectability with at least 80% power at the 5% significance level of within-stratum location differences corresponding to a "small" effect size (delta = 0.20), when the within-subject correlation in the responses across the Internet and telephone surveys is no lower than 0.60.

In Tables 2 and 3, we examine differences in the test-retest reliability of binary and ordinal variables across the four strata. Although not presented due to space limitations, the prevalence of binary and ordinal variables was calculated for both Internet- and telephone-administered measures. Prevalence differences

between the paired binary indicators contributed by each study subject were tested using the McNemar test of marginal homogeneity, as implemented in PROC FREQ of SAS v8.2 [74]. This test is equivalent to checking whether any disagreements that occur between the two methods of administration are entirely random and, hence, equally likely to be resolved in favor of either. It is noteworthy that its power is driven entirely by the number of subjects with discordant reports (N_D) rather than the total sample size. Effect sizes for the sign test have been defined by Cohen [75] as "small" for $g = 0.05$, "moderate" for $g = 0.15$, and "large" for $g = 0.25$, where g is the absolute difference from 50% in the proportion of discordant pairs that endorse the Internet over the interviewer-administered measure. Detectability with 80% power at the 5% significance level requires that N_D exceeds 140, 79, and 23, respectively. On this basis, "small" prevalence differences between Internet- and telephone-administered measures are detectable for all variables listed in Table 2, other than for the alcohol-related questions, for which only "moderate" differences can be detected.

In testing for prevalence differences between ordinal variables in Table 3, a latent variable model is assumed in which these variables can be construed as discretized versions on an underlying continuous variable. This model holds exactly for household income and appears quite reasonable for measuring health status. Because of skewness, the probit link associated with normal data in the latent scale was replaced by a log-log link for income and a complementary log-log link for health; both links can accommodate departures from symmetry and are related to the Gumbel distribution. In this setting, tests for prevalence differences between the Internet- and interviewer-administered measures translate into tests of location differences in the latent scale, implemented in PROC GENMOD of SAS v8.2 using generalized estimating equations, with a working exchangeable correlation matrix used to adjust for within-subject dependence in the paired ordinal measurements.

Shown in Tables 2 and 3 are the kappa coefficients [76], which measure the level of between-method agreement beyond that which can be ascribed to chance. Kappa coefficients have their range constrained by differences in prevalence between the dichotomous measures under investigation, and caution should be exercised in their interpretation when the associated sign test is significant [77]. In the absence of prevalence differences, standard cutoffs for measuring agreement have been established by Landis and Koch [78], which rate them as follows: 0.80-1.00 = almost perfect, 0.60-0.80 = substantial, 0.40-0.60 = moderate, 0.20-0.40 = fair, 0.00-0.20 = slight, and < 0.00 = poor. Confidence intervals for kappa coefficients have been calculated in Tables 2 and 3 using the profile variance method of Lee and Tu [79], which improves on the more common asymptotic normal approximation of Fleiss et al [80]. Extensions of kappa-type statistics to ordinal data have been proposed by Cohen [81] and require weights for the cells corresponding to partial agreement. Linearly decreasing weights of the form $1 - (i - j) / (k - 1)$ are employed, where i and j refer to the row and column scores and k is the number of categories. Health status has been rated on a 5-point scale, whereas household income is scored using the category midpoints for all categories other than the last one for which a sensitivity analysis was conducted by varying the midpoint from US $125,000 to $150,000. Finally, in Table 4 the internal consistency of several continuous scales is examined using Cronbach alpha coefficient [82], with the 95% CI obtained according to van Zyl et al [83].

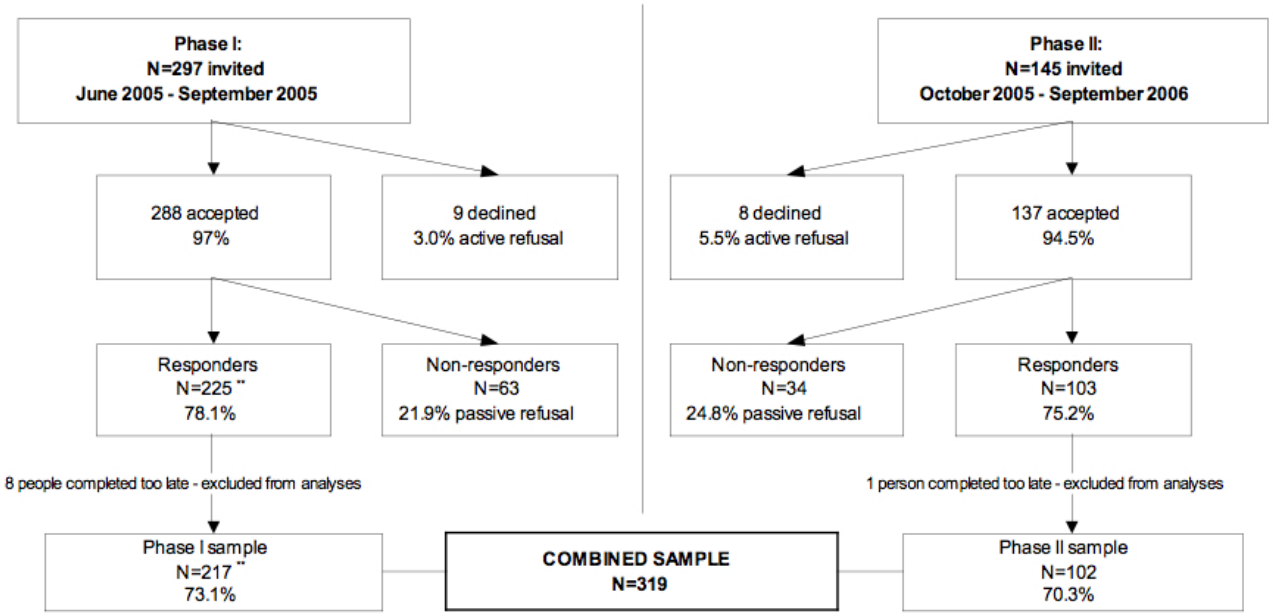
Results

Recruitment Results and Sample Characteristics

Details about enrollment are provided in Figure 1. During phase I of recruitment (June-September 2005), 297 individuals were invited to participate: 288 accepted (97%) and 217 (73.1%) completed the online survey within 1 week of their telephone assessment. Four individuals completed the online survey after data were pulled for the original analyses presented in our earlier study [46]. Thus, the sample size and response rates vary slightly from our original manuscript. During phase II of recruitment (October 2005 to September 2006), 145 individuals were invited to participate: 137 accepted (94.5%) and 102 (70.3%) completed the online survey within 1 week of their telephone assessment. The final sample size was 319. With regard to race/ethnicity, 52.4% were non-Hispanic White, 22.9% Black, 11.6% Hispanic, 7.8% Asian, 4.4% American Indian / Alaska Native, and 1% Native Hawaiian / Other Pacific Islander. About half (49.4%) of participants reported an annual household income of US $40,000 or less and 25.7% had a high school degree or less. The majority were women (61.4%), the average age of participants was 35.23 years (SD = 10.9; range 18-78), and participants smoked an average of 17.9 cigarettes per day (SD = 9.4; range 5-60).

There were no significant differences in participation rates between phase I and phase II of the study. The “active refusal rate” (ie, those who declined the initial invitation to participate) was 3.0% in phase I and 5.5% in phase II ($\chi^2_1 = 1.6$, $P = .20$). The “passive refusal rate” (ie, those who accepted the invitation to participate but did not complete the online survey) was 21.9% in phase I and 24.8% in phase II ($\chi^2_1 = 0.4$, $P = .49$).

Figure 1. Flowchart of enrollment in phase I and phase II



Means and Prevalence Data

For variables with negligible missingness (Smoking Temptations, PIQ, PSS, CES-D) the stratum-specific sample sizes were non-Hispanic Whites = 167, racial/ethnic minorities = 151, high income = 163, and low income = 151; for these variables, the study had at least 80% power at the 5% level of significance to detect stratum-specific effect sizes of delta = 0.22-0.23. For the alcohol variables, for which missingness rates were higher, the corresponding sample sizes were non-Hispanic Whites = 124, racial/ethnic minorities = 199, high income = 123, and low income = 99; for these variables the minimum detectable effect size rose to delta = 0.25-0.28. According to Cohen's [75] nomenclature, these are "small" effect sizes which, while likely to be statistically significant in our study, may be of less practical import than "moderate" effect sizes in the delta = 0.50 range.

As shown in Table 1, there was little systematic bias between the two survey methods as indicated by strong ICCs across all continuous variables. In examining mean differences, the only variable showing differences of “moderate” effect size between the Internet- and telephone-administered questionnaires is the Negative Affect subscale of Smoking Temptations, with the mean of the interviewer-administered measures 0.42-0.54 standard units higher than higher than the mean of the Internet-administered version across strata (all *P* values < .001). As a result, the total score of the Smoking Temptations scale shows an overall mean difference in the “small to moderate” range, with the two sample means 0.29-0.44 standard units apart (all *P* values < .001). Despite statistically significant differences for variables such as the PIQ total score and the CES-D measured among non-Hispanic Whites, the observed effect sizes were “small,” a result of the ample power our sample size affords for detecting within-subject differences in continuous outcomes.

Table 1. Internet–telephone reliabilities of continuous variables by race/ethnicity and income

	Racial/Ethnic Minority		Non-Hispanic White		Low Income		High Income	
	ICC	95% CI	ICC	95% CI	ICC	95% CI	ICC	95% CI
Smoking Temptations (total)	0.73	0.65-0.80	0.67	0.58-0.75	0.67	0.57-0.75	0.73	0.65-0.79
Positive Affect or Social Situations	0.70	0.61-0.77	0.65	0.55-0.73	0.66	0.56-0.74	0.69	0.59-0.76
Negative Affect Situations	0.67	0.57-0.75	0.66	0.56-0.74	0.69	0.60-0.77	0.63	0.54-0.72
Habitual or Craving Situations	0.69	0.60-0.77	0.70	0.61-0.77	0.70	0.62-0.78	0.69	0.60-0.76
PIQ (total)	0.88	0.84-0.91	0.90	0.87-0.93	0.89	0.85-0.92	0.90	0.87-0.93
Positive	0.75	0.67-0.81	0.85	0.80-0.88	0.75	0.68-0.82	0.85	0.80-0.89
Negative	0.87	0.83-0.91	0.91	0.88-0.94	0.89	0.85-0.92	0.91	0.87-0.93
PSS	0.77	0.70-0.83	0.74	0.66-0.80	0.79	0.73-0.85	0.71	0.62-0.78
CES-D	0.81	0.75-0.86	0.78	0.71-0.83	0.79	0.72-0.84	0.79	0.72-0.84
Alcohol Use								
Number of drinking days per week	0.94	0.91-0.96	0.95	0.93-0.96	0.96	0.93-0.97	0.93	0.90-0.95
Number of drinks on a typical day	0.91	0.87-0.94	0.83	0.76-0.88	0.93	0.89-0.95	0.77	0.69-0.83
Max number of drinks on a single occasion	0.92	0.87-0.95	0.93	0.91-0.95	0.97	0.95-0.98	0.91	0.87-0.93

Of the binary variables listed in Table 2, only two variables showed statistically significant differences in prevalence between the two survey methods. Across strata, the prevalence of self- reported smoking-related illness was 7%-13% higher when assessed over the phone (all P values < .02). Among non-Hispanic White and low-income participants, the use of pamphlets or booklets as quit aids was 6%-7% higher when assessed over the Internet (non-Hispanic White: 24.5% vs 17.9%, P = .02; low income: 25.2% vs 19.4%, P < .02). It should be noted that the post hoc power of the McNemar test in the present study is quite low for all but “large” effect sizes due to the small number of discordant pairs (N_D < 26 throughout). The ordinal variables listed in Table 2 (income, health status) showed no significant differences in prevalence under a latent variable model (all P values > .09).

Test-Retest Reliability and Internal Consistency Results

As seen in Table 1, test-retest reliability across modes of survey administration exceeded the minimal threshold of 70% for the majority of measures across strata. Reliabilities were very high (above 90%) for the alcohol use measures, with the exception of the number of drinks per typical day, for which reliabilities were lower for non-Hispanic Whites (0.83) and high-income subjects (0.77). For the PIQ, reliability was around 90% for both the total score and Negative Affect subscale, but dropped to 75% for the Positive Affect / Social subscale among racial/ethnic minority and low-income respondents. Reliability was moderately strong (in the 78%-81% range) for the CES-D and acceptable (in the 71%-79% range) for PSS. Results were least satisfactory for the individual subscales of the Smoking Temptations scale, none of which exceeded the 70% reliability threshold. Still, the overall scale (total score) was more reliable as would be expected from a composite of three correlated subscales, with its ICC exceeding the 70% threshold among racial/ethnic minority and high-income respondents.

In Table 2, almost perfect agreement between the two survey methods (kappa in the range of 0.80-1.00) was obtained across strata for 7 of the 15 binary variables assessing prior use of quit methods: nicotine patch, nicotine gum, nicotine inhaler, Zyban (bupropion), switching to chewing tobacco or snuff, acupuncture, and hypnosis. Use of the nicotine lozenge also showed near perfect agreement across all strata with the exception of racial/ethnic minority respondents, for which substantial agreement was obtained (kappa = 0.68). Three quit methods showed substantial degrees of agreement across all strata (kappa in the range of 0.60-0.80): use of pamphlet or booklet, group counseling, and telephone counseling. At least moderate agreement was obtained across strata for quitting cold turkey (kappa in the range of 0.56-0.71) and individual counseling (kappa in the range of 0.40-0.60). Agreement was poor to fair for Internet use (kappa in the range of 0.18-0.49). Reported use of nicotine spray as a quit method (which was infrequent among all respondents) showed poor agreement across surveys for racial/ethnic minority and low-income subjects, but moderate agreement among non-Hispanic White and high-income respondents. Although report of ever having a smoking-related illness showed substantial degrees of agreement across all respondents (kappa in the range of 0.65-0.71), this is a variable for which use of the kappa statistic may be inappropriate due to previously reported prevalence differences between the two survey methods [46]. As for the alcohol measures, all of them showed substantial to near perfect agreement across all four strata (all kappa values exceed 0.70).

Table 2. Internet–telephone reliabilities of binary variables by race/ethnicity and income

	Racial/Ethnic Minority		Non-Hispanic White		Low Income*		High Income*	
	Kappa†	95% CI	Kappa†	95% CI	Kappa†	95% CI	Kappa†	95% CI
Quit Methods (ever used)								
Cold turkey	0.69	0.52-0.86	0.58	0.42-0.73	0.71	0.55-0.88	0.56	0.40-0.72
Pamphlet or booklet	0.68‡	0.54-0.83	0.70	0.58-0.83	0.72‡	0.59-0.85	0.67	0.54-0.81
Individual counseling	0.57	0.31-0.84	0.48	0.13-0.84	0.61	0.34-0.87	0.44	0.09-0.78
Group counseling	0.69	0.51-0.88	0.80	0.63-0.97	0.68	0.48-0.87	0.81	0.65-0.97
Nicotine patch	0.96	0.91-1.00	0.92	0.85-0.98	0.95	0.90-1.00	0.93	0.87-0.98
Nicotine gum	0.91	0.84-0.98	0.92	0.86-0.98	0.90	0.82-0.97	0.93	0.88-0.99
Nicotine spray	−0.01	−0.02 to 0.00	0.66	0.05-1.00	−0.01	−0.02 to 0.00	0.49	−0.11 to 1.00
Nicotine lozenge	0.68	0.44-0.92	0.91	0.81-1.00	0.80	0.63-0.97	0.85	0.71-0.99
Nicotine inhaler	0.88	0.72-1.00	0.85	0.68-1.00	0.89	0.75-1.00	0.84	0.69-0.99
Zyban (bupropion)	0.93	0.86-1.00	0.94	0.88-0.99	0.90	0.83-0.98	0.96	0.91-1.00
Switch to chewing tobacco or snuff	0.85	0.68-1.00	0.86	0.73-1.00	0.89	0.6-1.00	0.82	0.65-0.99
Internet program	0.18	−0.09 to 0.44	0.49	0.23-0.75	0.37	0.11-0.63	0.32	0.02-0.62
Telephone counseling	0.88	0.72-1.00	0.65	0.37-0.94	0.74	0.53-0.96	0.83	0.59-1.00
Acupuncture	0.85	0.65-1.00	1.00	1.00-1.00	0.93	0.79-1.00	0.95	0.85-1.00
Hypnosis	0.96	0.88-1.00	0.96	0.90-1.00	0.91	0.81-1.00	1.00	1.00-1.00
Smoking-Related Illness								
Ever had smoking-related illness?	0.65‡	0.54-0.77	0.71‡	0.61-0.82	0.67‡	0.55-0.78	0.70‡	0.60-0.81
Alcohol Use								
Do you drink alcohol?	0.91	0.84-0.98	0.90	0.82-0.98	0.90	0.83-0.97	0.91	0.84-0.99
More to drink than meant to	0.75	0.63-0.88	0.91	0.84-0.99	0.84	0.73-0.95	0.84	0.74-0.94
Wanted/needed to cut down	0.77	0.63-0.90	0.84	0.72-0.96	0.72	0.57-0.88	0.88	0.78-0.98

*Income scored at category midpoint; US $125,000 used for last category.

†Weighted kappa using absolute difference between category scores to define a distance measure.

‡McNemar test for prevalence differences is significant.

In Table 3 we find almost perfect agreement for the income measure (weighted kappa values > 0.84) and substantial agreement for health status (weighted kappa values > 0.72). Results for the income measure were not dependent on whether the midpoint of the highest income category used to construct the weights was changed from US $125,000 to $150,000. Due to the informativeness of ordinal (as opposed to binary) measures, the confidence intervals are narrower, which indicates improved precision in the estimates.

Table 3. Internet–telephone reliabilities of ordinal variables by race/ethnicity and income

	Racial/Ethnic Minority		Non-Hispanic White		Low Income*		High Income*	
	Kappa†	95% CI	Kappa†	95% CI	Kappa†	95% CI	Kappa†	95% CI
Income†	0.87	0.82-0.93	0.94	0.90-0.97	0.84	0.78-0.91	0.92	0.88-0.97
Health status	0.73	0.66-0.81	0.72	0.64-0.81	0.72	0.64-0.80	0.74	0.66-0.83

*Income scored at category midpoint; US $125,000 used for last category.

†Weighted kappa using absolute difference between category scores to define a distance measure.

Finally, Table 4 reports the internal consistency of four scales of interest (total score only) and contrasts it across survey methods. For all four scales, the Internet-administered versions consistently have higher internal consistency than the interviewer-administered ones. Across all four strata, Cronbach alpha coefficients approach or exceed 80% for CES-D under both methods, are in the 70%-80% range for the PIQ and PSS, and only fall below 70% for the Smoking Temptations scale. Cross-survey comparisons show no statistically significant differences for PIQ, PSS, and Smoking Temptations, but are significant at the 5% level across all strata for the CES-D scale.

Although not reported in Table 4 due to space considerations, we also examined internal consistency of each of the three subscales of the Smoking Temptations Questionnaire within each of the four strata. The Negative Affect subscale maintained acceptable internal consistency levels across all four strata of interest, in the range of 77%-85% for Internet administration and 76%-78% for telephone administration. This was not the case for the Positive Affect / Social and Habit/Addictive scales, for which internal consistency levels never exceeded 60% in any of the four strata under both methods of administration. Full tables are available from the corresponding author.

Table 4. Internal consistency of measurement scales: Internet–telephone comparisons stratified by race/ethnicity and income

		Internet Administered		Telephone Administered	
	No. of Items*	Alpha†	95% CI	Alpha†	95% CI
Smoking Temptations					
Racial/ethnic minority	9	0.70	0.61-0.76	0.61	0.51-0.70
Non-Hispanic White	9	0.65	0.56-0.72	0.57	0.45-0.65
Low income	9	0.66	0.57-0.73	0.58	0.47-0.67
High income	9	0.68	0.60-0.75	0.60	0.50-0.69
PIQ					
Racial/ethnic minority	6	0.79	0.73-0.84	0.74	0.67-0.80
Non-Hispanic White	6	0.80	0.74-0.84	0.76	0.69-0.81
Low income	6	0.81	0.76-0.85	0.78	0.72-0.83
High income	6	0.79	0.73-0.83	0.72	0.65-0.78
PSS					
Racial/ethnic minority	4	0.75	0.68-0.81	0.75	0.68-0.81
Non-Hispanic White	4	0.77	0.71-0.82	0.71	0.62-0.77
Low income	4	0.75	0.67-0.80	0.72	0.63-0.78
High income	4	0.77	0.70-0.82	0.74	0.67-0.80
CES-D					
Racial/ethnic minority	10	0.86	0.82-0.89	0.82	0.77-0.86
Non-Hispanic White	10	0.86	0.82-0.89	0.79	0.73-0.83
Low income	10	0.85	0.81-0.88	0.79	0.73-0.83
High income	10	0.85	0.81-0.88	0.82	0.77-0.86

*Number of items in measurement scale.

†Cronbach alpha based on total score of unstandardized items for each scale.

Discussion

This study demonstrated that the psychometric properties of a broad range of measures commonly used in smoking cessation clinical trials are not different when administered via the Internet to racial/ethnic minority or low-income participants. Few studies to date have explicitly examined race/ethnicity and income with sufficient sample size and power to determine the degree of consistency between Internet- and telephone-administered questions, and none have examined these questions for cessation constructs. Therefore, these results provide new and largely reassuring information about measurement and method variance across two modes of administration in samples of participants who are of increasing importance to researchers involved in tobacco use behavior and cessation intervention research. Given the high smoking prevalence rates among racial/ethnic minority and low-income individuals, it is important to be able to reach and intervene in these target groups and to know that important data about key variables such as mediators, moderators, covariates, and outcomes can be collected using efficient modalities such as the Internet.

While the majority of measures were consistent across modes of administration, there were several statistically significant differences in means and prevalence. In general, these differences have minimal clinical significance as they were small in magnitude; however, such differences highlight the

importance of pilot testing items with the target population to ensure adequate comprehension of questions as well as response formats. Items that require clarification by telephone or that yield different means or prevalence when administered via the Internet may require more detailed instructions or specific illustrative examples to assist research participants. The Internet is more similar to a paper-and-pencil test than an interview during which prompts and clarifications can be made. Equivalent forms of questions need to be tested to ensure all items and scales are consistent across modes of delivery whenever possible. Detecting differences can help improve the reliability, validity, and equivalency of measures across modalities. Empirical data of the kind collected in this study can provide valuable information to researchers about possible sources of error variance or systematic measurement bias.

The majority of the test-retest reliability coefficients fell above the minimum threshold of 0.70, indicating substantial to strong agreement between survey methods. Two exceptions noted were in the quit methods measure in items that assessed use of nicotine spray and Internet cessation websites in previous quit attempts. In general, these two findings should be interpreted with caution given that overall prevalence of both quit methods was very low in both the phone and Internet surveys (< 10%) and that both point and interval estimates of kappa are extremely sensitive to small changes in cell counts. However, we also know from analyses of follow-up data for the parent study that some participants continued looking for cessation assistance on the Internet following randomization. When reporting on use of smoking cessation websites in the Internet survey, participants may have included their use of cessation websites following the baseline telephone assessment. The take-home message here is that it is critical to examine the time frame referenced in a reliability study to ensure that the wording of questions does not artificially inflate or deflate the concordance of responses. Participants were asked to indicate whether they had "ever used" a variety of quit methods. It is reasonable to consider that participants in the smoking cessation parent trial began trying various methods of quitting immediately following enrollment and referenced those methods in the Internet survey (administered following randomization) but not in the baseline telephone survey.

Internal consistency across items was good for all scales examined, with the exception of the Smoking Temptations scale. Across strata, Cronbach alpha coefficients did not exceed the threshold of 0.70 for either the total scale score or any of the subscales. These findings are consistent with our previously published study [46] and with work by Ward (personal communication, RM Ward, October 2007) in which Cronbach alpha coefficients were as follows: Negative Affect = 0.765, Habit/Addictive = 0.579, and Positive Affect / Social = 0.573. Given the poor performance of the Habit/Addictive and Positive Affect / Social subscales, it is not surprising that the internal consistency of the overall Smoking Temptations scale failed to exceed the 70% threshold across strata and mode of administration in the present study. Given these findings, further refinement of the Short-Form of this measure is called for, especially since the availability of psychometrically sound, brief assessment instruments is critical to minimize response burden in Web-based smoking cessation research trials.

Results should be considered in the context of several limitations. First, it is possible that the higher internal consistency seen in Internet-administered measures was due to learning effects since Internet measures were always administered several days after the telephone interview. Counterbalancing the order of administration would address this limitation and should be considered in future studies. Second, it is possible that the internal consistency may have been artificially inflated due to memory effects associated with the relatively short (ie, 2-7 days) time frame between measurement points. There is often a 2-4 week gap between repeat administrations of the same scales for test-retest reliability studies. Given the dynamic nature of many of the constructs we assessed—especially in the context of a cessation trial—this shorter time frame was necessary so as not to artificially deflate internal consistency due to expected changes in knowledge, attitudes, and beliefs. Third, some may question our use of a cutoff of US $40,000 for our "low income" stratum. The median household income in the United States 2006 was US $48,201 [84], meaning that half the US population fell below this threshold. In addition, Internet use is more common in households with higher levels of income (93% for ≥ US $75,000 vs 49% for < US $30,000 [14]). Therefore, we believe that US $40,000 or less is a reasonable cutoff for lower income Internet users. Finally, our use of race/ethnicity as a categorization variable was to explore in a preliminary fashion whether there are differences by culture or context in the psychometric properties of Internet-administered measures. A limitation of this approach is that the group of racial/ethnic minority participants is likely still quite heterogeneous with regard to race, ethnicity, and other variables that may influence survey response patterns. Future studies should move beyond race and ethnicity to investigate the specific factors that may link race/ethnicity to measurement issues such as health literacy, technology access and familiarity, and other cultural factors.

In conclusion, the present study replicated findings from an earlier study demonstrating adequate internal consistency and test-retest reliability of a broad range of measures commonly used in smoking cessation clinical trials. In addition, this study extended these findings by examining measures among racial/ethnic minorities and individuals with lower levels of household income. This knowledge adds to the confidence of conducting Web-based research and strengthens the scientific rigor of collecting information via the Internet on racial/ethnic minority and low-income subgroups. This study also revealed a few areas where measurement scales did not perform as well as expected. These findings underscore the importance of explicitly testing consistency among subgroups with sufficient statistical power in order to test empirically the equivalence of measures and to identify measures that require more work to improve their performance in specific subgroups.

Acknowledgments

This research was supported by the National Cancer Institute (R01CA104836). We acknowledge and thank Crystal Davis, MPH, Pearl Zakroysky, and Arielle Baskin-Sommers for their assistance on this project. QuitNet is a commercial website owned by Healthways QuitNet, LLC.

Conflicts of Interest

None declared.

References

1. ; Centers for Disease Control and Prevention (CDC). Tobacco use among adults--United States, 2005. MMWR Morb Mortal Wkly Rep 2006 Oct 27;55(42):1145-1148 [FREE Full text] [Medline: 17065979]
2. Burns DM, Anderson CM, Johnson M, Major JM, Biener L, Vaughn J, et al. Cessation and Cessation Measures among Adult Daily Smokers: National and State-Specific Data. In: Population Based Smoking Cessation: Proceedings of a Conference on What Works to Influence Cessation in the General Population. Smoking and Tobacco Control Monograph No. 12. Washington, DC: US Department of Health and Human Services; 2000. Nov:25-98.
3. Levy DT, Romano E, Mumford E. The relationship of smoking cessation to sociodemographic characteristics, smoking intensity, and tobacco control policies. Nicotine Tob Res 2005 Jun;7(3):387-396. [Medline: 16085506] [doi: 10.1080/14622200500125443]
4. ; Centers for Disease Control and Prevention (CDC). Annual smoking-attributable mortality, years of potential life lost, and economic costs--United States, 1995-1999. MMWR Morb Mortal Wkly Rep 2002 Apr 12;51(14):300-303 [FREE Full text] [Medline: 12002168]
5. ; Centers for Disease Control and Prevention (CDC). Annual smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 1997-2001. MMWR Morb Mortal Wkly Rep 2005 Jul 1;54(25):625-628 [FREE Full text] [Medline: 15988406]
6. ; Centers for Disease Control and Prevention (CDC). Tobacco Use Among US Racial/Ethnic Minority Groups–African Americans, American Indians and Alaska Natives, Asian Americans and Pacific Islanders, and Hispanics: A Report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services; 1998. URL: http://www.cdc.gov/tobacco/data_statistics/sgr/sgr_1998/00_pdfs/intro.pdf [accessed 2008 Feb 28] [WebCite Cache ID 5VPhoyR42]
7. ; US Department of Health and Human Services. Expanding the Reach and Impact of Consumer E Health Tools. Washington, DC: US Department of Health and Human Services, Office of Disease Prevention and Health Promotion; 2006. URL: http://www.health.gov/communication/ehealth/ehealthTools/default.htm [accessed 2008 Feb 28] [WebCite Cache ID 5VPhuQHyv]
8. Gibbons MC. A historical overview of health disparities and the potential of eHealth solutions. J Med Internet Res 2005;7(5):e50 [FREE Full text] [Medline: 16403714] [doi: 10.2196/jmir.7.5.e50]
9. Eng TR. Population health technologies: emerging innovations for the health of the public. Am J Prev Med 2004 Apr;26(3):237-242. [Medline: 15026105] [doi: 10.1016/j.amepre.2003.12.004]
10. Ahern DK. Challenges and opportunities of eHealth research. Am J Prev Med 2007 May;32(5 Suppl):S75-S82. [Medline: 17466822] [doi: 10.1016/j.amepre.2007.01.016]
11. ; US Department of Health and Human Services. Healthy People 2010: Understanding and Improving Health and Objectives for Improving Health. 2nd edition. Washington, DC: US Government Printing Office; 2000.
12. Graham AL, Abrams DB. Reducing the cancer burden of lifestyle factors: opportunities and challenges of the Internet. J Med Internet Res 2005 Jul 1;7(3):e26 [FREE Full text] [Medline: 15998617] [doi: 10.2196/jmir.7.3.e26]
13. Madden M. Internet penetration and impact. Washington, DC: Pew Internet & American Life Project; 2006. URL: http://www.pewinternet.org/PPF/r/182/report_display.asp [accessed 2008 Feb 28] [WebCite Cache ID 5VPhzLtVb]
14. Fox S. Digital divisions. Washington, DC: Pew Internet & American Life Project; 2005. URL: http://www.pewinternet.org/PPF/r/165/report_display.asp [accessed 2008 Feb 28] [WebCite Cache ID 5VPi2ZDVC]
15. ; US Department of Commerce. A Nation Online: How Americans Are Expanding Their Use of the Internet. Washington, DC: US Department of Commerce, Economics and Statistics Administration, National Telecommunications and Information Administration; 2002. URL: http://www.ntia.doc.gov/ntiahome/dn/nationonline_020502.htm [accessed 2008 Mar 3] [WebCite Cache ID 5VPi6OQoL]
16. Fox S. Health information online. Washington, DC: Pew Internet & American Life Project; 2005. URL: http://www.pewinternet.org/PPF/r/156/report_display.asp [accessed 2008 Mar 3] [WebCite Cache ID 5VPi9ZL96]
17. Fox S, Fallows D. Internet health resources: health searches and email have become more commonplace, but there is room for improvement in searches and overall Internet access. Washington, DC: Pew Internet & American Life Project; 2003. URL: http://www.pewinternet.org/report_display.asp?r=95 [accessed 2008 Mar 3] [WebCite Cache ID 5VPiCXnBG]
18. Bock B. Web-Assisted Tobacco Interventions (WATI) Initiative: Where we are, where we're going, and some challenges along the way. Paper presented at: 13th Annual Meeting of the Society for Research on Nicotine and Tobacco; February 21-24, 2007; Austin, TX. 2007.

19. Lenert L, Muñoz RF, Perez JE, Bansod A. Automated e-mail messaging as a tool for improving quit rates in an internet smoking cessation intervention. J Am Med Inform Assoc 2004;11(4):235-240 [FREE Full text] [Medline: 15064291] [doi: 10.1197/jamia.M1464]
20. Lenert L, Muñoz RF, Stoddard J, Delucchi K, Bansod A, Skoczen S, et al. Design and pilot evaluation of an internet smoking cessation program. J Am Med Inform Assoc 2003;10(1):16-20 [FREE Full text] [Medline: 12509354] [doi: 10.1197/jamia.M1128]
21. Feil EG, Noell J, Lichtenstein E, Boles SM, McKay HG. Evaluation of an Internet-based smoking cessation program: lessons learned from a pilot study. Nicotine Tob Res 2003 Apr;5(2):189-194. [Medline: 12745491] [doi: 10.1080/1462220031000073694]
22. Cobb NK, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. Nicotine Tob Res 2005 Apr;7(2):207-216. [Medline: 16036277] [doi: 10.1080/14622200500055319]
23. Cheh JA, Ribisl KM, Wildemuth BM. An assessment of the quality and usability of smoking cessation information on the Internet. Health Promot Pract 2003 Jul;4(3):278-287. [Medline: 14610998] [doi: 10.1177/1524839903004003012]
24. Bock B, Graham A, Sciamanna C, Krishnamoorthy J, Whiteley J, Carmona-Barros R, et al. Smoking cessation treatment on the Internet: content, quality, and usability. Nicotine Tob Res 2004 Apr;6(2):207-219. [Medline: 15203794] [doi: 10.1080/14622200410001676332]
25. Japuntich SJ, Zehner ME, Smith SS, Jorenby DE, Valdez JA, Fiore MC, et al. Smoking cessation via the internet: a randomized clinical trial of an internet intervention as adjuvant treatment in a smoking cessation intervention. Nicotine Tob Res 2006 Dec;8 Suppl 1(1):S59-S67. [Medline: 17491172] [doi: 10.1080/14622200601047900]
26. Strecher VJ, Shiffman S, West R. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. Addiction 2005 May;100(5):682-688. [Medline: 15847626] [doi: 10.1111/j.1360-0443.2005.01093.x]
27. Strecher VJ, Shiffman S, West R. Moderators and mediators of a web-based computer-tailored smoking cessation program among nicotine patch users. Nicotine Tob Res 2006 Dec;8 Suppl 1(1):S95-101. [Medline: 17491176] [doi: 10.1080/14622200601039444]
28. Graham AL, Bock BC, Cobb NK, Niaura R, Abrams DB. Characteristics of smokers reached and recruited to an internet smoking cessation trial: a case of denominators. Nicotine Tob Res 2006 Dec;8 Suppl 1(1):S43-S48. [Medline: 17491170] [doi: 10.1080/14622200601042521]
29. Swartz LHG, Noell JW, Schroeder SW, Ary DV. A randomised control study of a fully automated internet based smoking cessation programme. Tob Control 2006 Feb;15(1):7-12 [FREE Full text] [Medline: 16436397] [doi: 10.1136/tc.2003.006189]
30. Etter JF. Comparing the efficacy of two Internet-based, computer-tailored smoking cessation programs: a randomized trial. J Med Internet Res 2005;7(1):e2 [FREE Full text] [Medline: 15829474] [doi: 10.2196/jmir.7.1.e2]
31. Pike KJ, Rabius V, McAlister A, Geiger A. American Cancer Society's QuitLink: randomized trial of Internet assistance. Nicotine Tob Res 2007 Mar;9(3):415-420. [Medline: 17365773] [doi: 10.1080/14622200701188877]
32. Etter JF, le Houezec J, Landfeldt B. Impact of messages on concomitant use of nicotine replacement therapy and cigarettes: a randomized trial on the Internet. Addiction 2003 Jul;98(7):941-950. [Medline: 12814500] [doi: 10.1046/j.1360-0443.2003.00406.x]
33. ; NIH State-of-the-Science Panel. National Institutes of Health State-of-the-Science conference statement: tobacco use: prevention, cessation, and control. Ann Intern Med 2006 Dec 5;145(11):839-844 [FREE Full text] [Medline: 16954353]
34. ; Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
35. ; Institute of Medicine, Committee on the National Quality Report on Health Care Delivery. Envisioning the National Health Care Quality Report. Washington, DC: National Academies Press; 2001.
36. ; Institute of Medicine, Committee on Communication for Behavior Change in the 21st Century. Speaking of Health: Assessing Health Communication Strategies for Diverse Populations. Washington, DC: National Academies Press; 2002.
37. ; Institute of Medicine. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington, DC: National Academies Press; 2002.
38. Muñoz RF, Lenert LL, Delucchi K, Stoddard J, Perez JE, Penilla C, et al. Toward evidence-based Internet interventions: A Spanish/English Web site for international smoking cessation trials. Nicotine Tob Res 2006 Feb;8(1):77-87. [Medline: 16497602] [doi: 10.1080/14622200500431940]
39. Byrne BM, Watkins D. The issue of measurement invariance revisited. J Cross Cult Psychol 2003;34(2):155-175. [doi: 10.1177/0022022102250225]
40. Myers MB, Calantone RJ, Page TJJ, Taylor CR. An application of multiple-group causal models in assessing cross-cultural measurement equivalence. J Int Mark 2000;8(4):108-121. [doi: 10.1509/jimk.8.4.108.19790]
41. Ramírez M, Ford ME, Stewart AL, Teresi JA. Measurement issues in health disparities research. Health Serv Res 2005 Oct;40(5 Pt 2):1640-1657 [FREE Full text] [Medline: 16179000] [doi: 10.1111/j.1475-6773.2005.00450.x]
42. Boeckner LS, Pullen CH, Walker SN, Abbott GWA, Block T. Use and reliability of the World Wide Web version of the Block Health Habits and History Questionnaire with older rural women. J Nutr Educ Behav 2002;34 Suppl 1(Suppl 1):S20-S24. [Medline: 12047825] [doi: 10.1016/S1499-4046(06)60307-2]

43. Bressani RV, Downs AC. Youth independent living assessment: testing the equivalence of Web and paper/pencil versions of the Ansell-Casey Life Skills Assessment. Comput Human Behav 2002;18(4):453. [doi: 10.1016/S0747-5632(01)00053-X]
44. Ritter P, Lorig K, Laurent D, Matthews K. Internet versus mailed questionnaires: a randomized comparison. J Med Internet Res 2004 Sep 15;6(3):e29 [FREE Full text] [Medline: 15471755] [doi: 10.2196/jmir.6.3.e29]
45. Vallejo MA, Jordán CM, Díaz MI, Comeche MI, Ortega J. Psychological assessment via the internet: a reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). J Med Internet Res 2007;9(1):e2 [FREE Full text] [Medline: 17478411] [doi: 10.2196/jmir.9.1.e2]
46. Graham AL, Papandonatos GD, Bock BC, Cobb NK, Baskin-Sommers A, Niaura R, et al. Internet- vs. telephone-administered questionnaires in a randomized trial of smoking cessation. Nicotine Tob Res 2006 Dec;8 Suppl 1(1):S49-S57. [Medline: 17491171] [doi: 10.1080/14622200601045367]
47. ; American Psychological Association (APA). Ethical principles of psychologists and code of conduct. Am Psychol 2002;57(12):1060-1073. [Medline: 12613157] [doi: 10.1037/0003-066X.57.12.1060]
48. Okazaki S, Sue S. Methodological issues in assessment research with ethnic minorities. In: In: Kazdin AE, editor. Methodological Issues and Strategies in Clinical Research. 2nd edition. Washington, DC: American Psychological Association; 1998:263-281.
49. ; Office of Management and Budget. Federal Register Notice. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. 1997 Oct 30. OMB. URL: http://www.whitehouse.gov/omb/fedreg/1997standards.html [accessed 2008 Mar 3] [WebCite Cache ID 5VPSgdXNz]
50. Velicer WF, Diclemente CC, Rossi JS, Prochaska JO. Relapse situations and self-efficacy: an integrative model. Addict Behav 1990;15(3):271-283. [Medline: 2378287] [doi: 10.1016/0306-4603(90)90070-E]
51. Coppotelli HC, Orleans CT. Partner support and other determinants of smoking cessation maintenance among women. J Consult Clin Psychol 1985 Aug;53(4):455-460. [Medline: 4031200] [doi: 10.1037/0022-006X.53.4.455]
52. Mermelstein R, Lichtenstein E, McIntyre K. Partner support and relapse in smoking-cessation programs. J Consult Clin Psychol 1983 Jun;51(3):465-466. [Medline: 6863711] [doi: 10.1037/0022-006X.51.3.465]
53. Cohen S, Lichtenstein E. Partner behaviors that support quitting smoking. J Consult Clin Psychol 1990 Jun;58(3):304-309. [Medline: 2365893] [doi: 10.1037/0022-006X.58.3.304]
54. Roski J, Schmid LA, Lando HA. Long-term associations of helpful and harmful spousal behaviors with smoking cessation. Addict Behav 1996;21(2):173-185. [Medline: 8730519] [doi: 10.1016/0306-4603(95)00047-X]
55. McMahon SD, Jason LA. Social support in a worksite smoking intervention. A test of theoretical models. Behav Modif 2000 Apr;24(2):184-201. [Medline: 10804679] [doi: 10.1177/0145445500242002]
56. Gruder CL, Mermelstein RJ, Kirkendol S, Hedeker D, Wong SC, Schreckengost J, et al. Effects of social support and relapse prevention training as adjuncts to a televised smoking-cessation intervention. J Consult Clin Psychol 1993 Feb;61(1):113-120. [Medline: 8450096] [doi: 10.1037/0022-006X.61.1.113]
57. Cohen S, Lichtenstein E, Prochaska JO, Rossi JS, Gritz ER, Carr CR, et al. Debunking myths about self-quitting. Evidence from 10 prospective studies of persons who attempt to quit smoking by themselves. Am Psychol 1989 Nov;44(11):1355-1365. [Medline: 2589730] [doi: 10.1037/0003-066X.44.11.1355]
58. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417] [doi: 10.2307/2136404]
59. Cohen S, Williamson GM. Perceived stress in a probability sample of the United States. In: In: Spacapan S, Oskamp S, editors. The Social Psychology of Health. Newbury Park, CA: Sage; 1988:31-67.
60. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 1994;10(2):77-84. [Medline: 8037935]
61. Anda RF, Williamson DF, Escobedo LG, Mast EE, Giovino GA, Remington PL. Depression and the dynamics of smoking. A national perspective. JAMA 1990 Sep 26;264(12):1541-1545. [Medline: 2395193] [doi: 10.1001/jama.264.12.1541]
62. Niaura R, Britt DM, Borrelli B, Shadel WG, Abrams DB, Goldstein MG. History and symptoms of depression among smokers during a self-initiated quit attempt. Nicotine Tob Res 1999 Sep;1(3):251-257. [Medline: 11072422] [doi: 10.1080/14622299050011371]
63. Pomerleau CS, Zucker AN, Stewart AJ. Patterns of depressive symptomatology in women smokers, ex-smokers, and never-smokers. Addict Behav 2003 Apr;28(3):575-582. [Medline: 12628628] [doi: 10.1016/S0306-4603(01)00257-X]
64. Lerman C, Audrain J, Orleans CT, Boyd R, Gold K, Main D, et al. Investigation of mechanisms linking depressed mood to nicotine dependence. Addict Behav 1996;21(1):9-19. [Medline: 8729703] [doi: 10.1016/0306-4603(95)00032-1]
65. Zhu SH, Valbø A. Depression and smoking during pregnancy. Addict Behav 2002;27(4):649-658. [Medline: 12188598] [doi: 10.1016/S0306-4603(01)00199-X]
66. Munafò MR, Hitsman B, Rende R, Metcalfe C, Niaura R. Effects of progression to cigarette smoking on depressed mood in adolescents: evidence from the National Longitudinal Study of Adolescent Health. Addiction 2008 Jan;103(1):162-171. [Medline: 18031565] [doi: 10.1111/j.1360-0443.2007.02052.x]

67. Benjet C, Wagner FA, Borges GG, Medina-Mora ME. The relationship of tobacco smoking with depressive symptomatology in the Third Mexican National Addictions Survey. Psychol Med 2004 Jul;34(5):881-888. [Medline: 15500308] [doi: 10.1017/S0033291703001600]
68. McClure JB, Wetter DW, de Moor C, Cinciripini PM, Gritz ER. The relation between alcohol consumption and smoking abstinence: results from the Working Well Trial. Addict Behav 2002;27(3):367-379. [Medline: 12125663] [doi: 10.1016/S0306-4603(01)00177-0]
69. Shiffman S, Hickcox M, Paty JA, Gnys M, Kassel JD, Richards TJ. Progression from a smoking lapse to relapse: prediction from abstinence violation effects, nicotine dependence, and lapse characteristics. J Consult Clin Psychol 1996 Oct;64(5):993-1002. [Medline: 8916628] [doi: 10.1037/0022-006X.64.5.993]
70. ; Centers for Disease Control and Prevention. 2002 Behavioral Risk Factor Surveillance System. 2001 Dec 27. URL: http://www.cdc.gov/brfss/questionnaires/pdf-ques/2002brfss.pdf [accessed 2008 Mar 3] [WebCite Cache ID 5VPSPH9F7]
71. Brown RL, Leonard T, Saunders LA, Papasouliotis O. A two-item conjoint screen for alcohol and other drug problems. J Am Board Fam Pract 2001;14(2):95-106 [FREE Full text] [Medline: 11314930]
72. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914] [doi: 10.1097/00005650-199206000-00002]
73. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420-428. [doi: 10.1037/0033-2909.86.2.420]
74. ; SAS/STAT Software. Changes and Enhancements. Cary, NC: SAS Institute, Inc; 2001.
75. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum; 1988.
76. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37-46. [doi: 10.1177/001316446002000104]
77. Cook RJ. Kappa and its dependence on marginal rates. In: Armitage P, Colton T, editors. The Encyclopedia of Biostatistics. New York: Wiley; 1998:2166-2168.
78. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571] [doi: 10.2307/2529310]
79. Lee JJ, Tu ZN. A better confidence for kappa on measuring agreement between two raters with binary outcomes. J Comput Graph Stat 1994;3(3):301-321. [doi: 10.2307/1390914]
80. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. Psychol Bull 1969;72(5):323-327. [doi: 10.1037/h0028106]
81. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70(4):213-220. [doi: 10.1037/h0026256]
82. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16(3):297-333. [doi: 10.1007/BF02310555]
83. van Zyl JM, Neudecker H, Nel DG. On the distribution of the maximum likelihood estimator of Cronbach's alpha. Psychometrika 2000;65(3):271-280. [doi: 10.1007/BF02296146]
84. DeNavas-Walt C, Proctor BD, Lee CH. US Census Bureau. Income, Poverty, and Health Insurance Coverage in the United States: 2006 Aug. census.gov. URL: http://www.census.gov/prod/2006pubs/p60-231.pdf [accessed 2008 Mar 3] [WebCite Cache ID 5VPgltUHM]

submitted 28.10.07; peer-reviewed by JF Etter, A Palepu; comments to author 11.01.08; revised version received 07.02.08; accepted 20.02.08; published 26.03.08.

Please cite as:
Graham AL, Papandonatos GD
Reliability of Internet- Versus Telephone-Administered Questionnaires in a Diverse Sample of Smokers
J Med Internet Res 2008;10(1):e8
URL: http://www.jmir.org/2008/1/e8/
doi:10.2196/jmir.987
PMID:18364345

Original Paper

Impact of Web Searching and Social Feedback on Consumer Decision Making: A Prospective Online Experiment

Annie YS Lau, PhD; Enrico W Coiera, MB, BS, PhD

[1]

Corresponding Author:
Enrico W Coiera, MB, BS, PhD
Centre for Health Informatics
University of New South Wales
Sydney, NSW 2052
Australia
Phone: +61 2 9385 9026
Fax: +61 2 9385 9006
Email: e.coiera@unsw.edu.au

Abstract

Background: The World Wide Web has increasingly become an important source of information in health care consumer decision making. However, little is known about whether searching online resources actually improves consumers' understanding of health issues.

Objectives: The aim was to study whether searching on the World Wide Web improves consumers' accuracy in answering health questions and whether consumers' understanding of health issues is subject to further change under social feedback.

Methods: This was a pre/post prospective online study. A convenience sample of 227 undergraduate students was recruited from the population of the University of New South Wales. Subjects used a search engine that retrieved online documents from PubMed, MedlinePlus, and HealthInsite and answered a set of six questions (before and after use of the search engine) designed for health care consumers. They were then presented with feedback consisting of a summary of the post-search answers provided by previous subjects for the same questions and were asked to answer the questions again.

Results: There was an improvement in the percentage of correct answers after searching (pre-search 61.2% vs post-search 82.0%, P <.001) and after feedback with other subjects' answers (pre-feedback 82.0% vs post-feedback 85.3%, P =.051).The proportion of subjects with highly confident correct answers (ie, confident or very confident) and the proportion with highly confident incorrect answers significantly increased after searching (correct pre-search 61.6% vs correct post-search 95.5%, P <.001; incorrect pre-search 55.3% vs incorrect post-search 82.0%, P <.001). Subjects who were not as confident in their post-search answers were 28.5% more likely than those who were confident or very confident to change their answer after feedback with other subjects' post-search answers (χ^2_1= 66.65, P <.001).

Conclusions: Searching across quality health information sources on the Web can improve consumers' accuracy in answering health questions. However, a consumer's confidence in an answer is not a good indicator of the answer being correct. Consumers who are not confident in their answers after searching are more likely to be influenced to change their views when provided with feedback from other consumers.

(J Med Internet Res 2008;10(1):e2) doi:10.2196/jmir.963

KEYWORDS

Health care consumer; information searching; decision making; social feedback; Internet; accuracy; confidence

Introduction

The World Wide Web is now recognized as an important source of information in supporting the practice of evidence-based medicine [1] and consumer health care decision making [2]. While much research focuses on the impact of information retrieval on clinical decision making, there has been little examination of how online searching influences the way consumers make health-related decisions.

Many studies have examined the quality of online health care consumer information [3], the tools and initiatives developed to promote health literacy [4], as well as the characteristics of websites and search engines that influence the way consumers perceive and utilize information [5,6]. Of particular relevance

to understanding the way consumers use online health-related information, past studies have examined consumers' familiarity with health vocabulary [7], their information appraisal [8] and search query reformulation skills [9], the way they perceive and assess Web-based health information [10-13], the types of online sources they trust [14], the patterns of use and barriers experienced while using online resources [15], and how access to online information influences the way they interact with health care professionals [16-18].

Studies have also shown that people are an important source of influence among consumers with a health-related concern. In a randomized controlled trial conducted by Lorig et al, patients with back pain who had access to an email discussion group demonstrated greater improvement in pain and made less physician visits than those without access [19]. Patients with breast cancer participating in electronic support groups are reported to have reduced rates of depression and lessened reactions to pain [20].

Little, however, is known about whether consumers are actually able to improve their understanding of health issues after searching the Web. In addition, little is known about the extent to which social feedback affects the way consumers develop their understanding of health issues. This prospective experiment tests the following hypotheses: (1) consumers can improve their accuracy in answering health care questions after searching tested online resources, and (2) consumers'answers to health care questions are influenced by feedback with other consumers' answers.

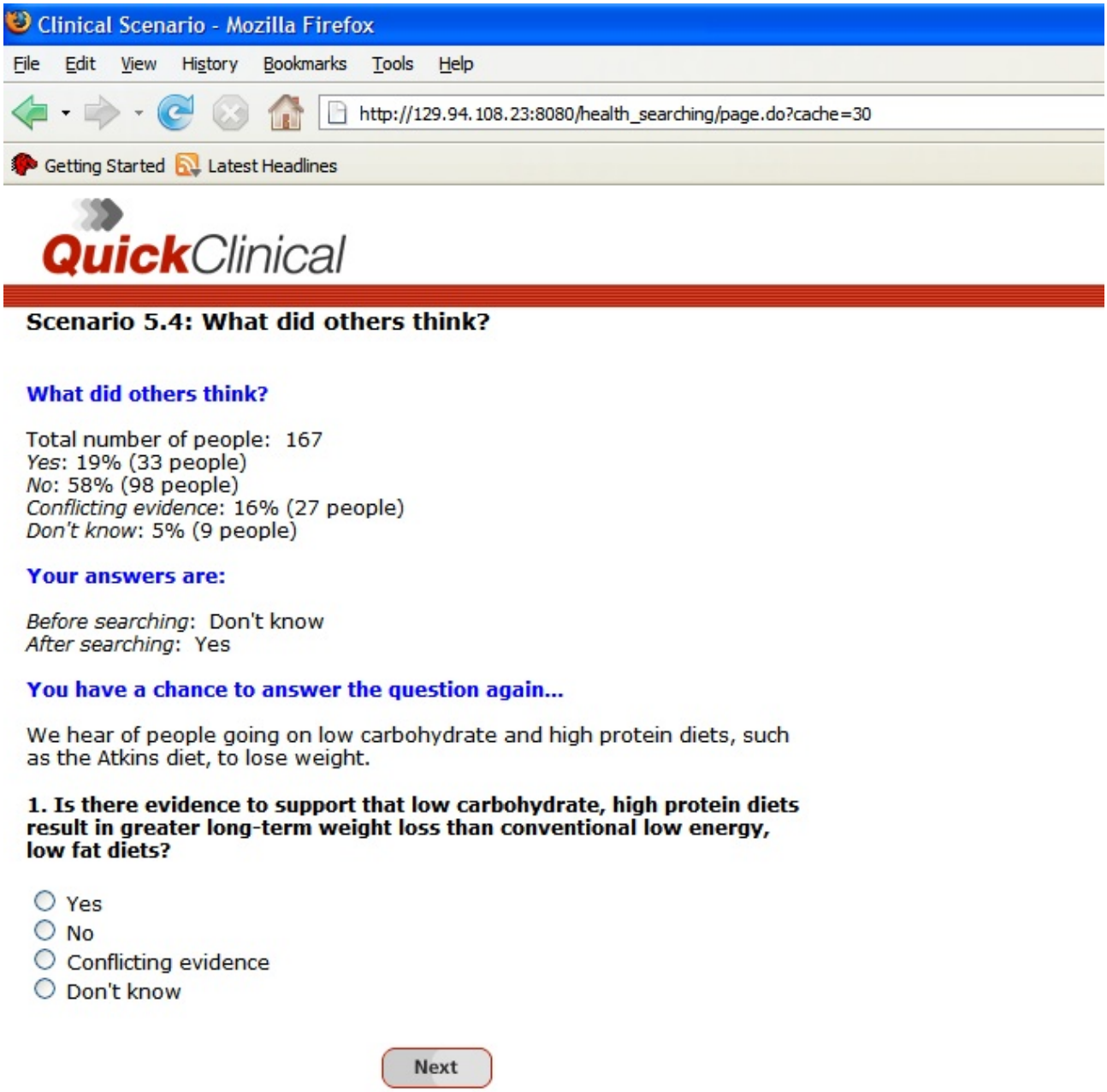
Methods

Study Design

A convenience sample of 227 undergraduate students was recruited from the University of New South Wales (UNSW). Subjects were asked to use a specific online search engine to answer six consumer health questions. People with Internet access who had previously used an online search engine were recruited by announcements via student email lists, posters, leaflets, weekly student magazines, and a UNSW research news website. Upon completion of the study, subjects were entered into a draw for one of 100 movie tickets. Ethics approval was obtained from the Human Research Ethics Advisory Panel at UNSW.

A pre/post protocol was used in this study. Subjects recorded their pre- and post-search answers to each question and their confidence in these answers. After answering each question post-search, subjects were presented with a summary of the post-search answers provided by previous subjects and were asked to answer the question again (Figure 1).

Figure 1. Screen capture of feedback provided to subjects after answering a question post-search



Each question and the expected correct answer are shown in Table 1. All scenario questions were randomly allocated. There were four options to answer a question: "yes," "no," "conflicting evidence," and "don't know." Confidence was measured by a 4-point Likert scale from "very confident" to "not confident." The questions ranged in difficulty and topic in order to cover a spectrum of health care consumer questions. They were developed in consultation with a general practitioner and two academics from the School of Public Health and Community Medicine at UNSW. Agreement was reached on the "correct" answer and the location of the best evidence sources for each question. A pilot test with three members of the general public tested the questions for interest and readability. Two additional pilots of five people each were conducted to confirm that it was possible to locate documentary evidence required to answer the questions correctly.

The search engine retrieved documents from tested resources known to have high relevance in answering health-related questions [21]. These resources are PubMed [22], MedlinePlus [23], and HealthInsite [24]. Overall, subjects were advised to spend about 10 minutes for each question and to use only the provided search system to answer the questions. To prevent subjects from visiting external websites during the experiment, the navigational bar on the Web browser was hidden once the subject logged on to the study website.

Table 1. Case scenarios and questions presented to subjects. A random selection of six cases was presented to each subject in the study.

Case Scenario and Question (Scenario Name)	Expected Correct Answer
1. We hear of people going on low carbohydrate and high protein diets, such as the Atkins diet, to lose weight. *Is there evidence to support that low carbohydrate, high protein diets result in greater long-term weight loss than conventional low energy, low fat diets?* (Diet)	No
2. You can catch infectious diseases such as the flu from inhaling the air into which others have sneezed or coughed, sharing a straw, or eating off someone else's fork. The reason is because certain germs reside in saliva, as well as in other bodily fluids. Hepatitis B is an infectious disease. *Can you catch Hepatitis B from kissing on the cheek?* (Hepatitis B)	No
3. After having a few alcoholic drinks, we depend on our liver to reduce the Blood Alcohol Concentration (BAC). Drinking coffee, eating, vomiting, sleeping, or having a shower will not help reduce your BAC. *Are there different recommendations regarding safe alcohol consumption for males and females?* (Alcohol)	Yes
4. Sudden infant death syndrome (SIDS), also known as "cot death," is the unexpected death of a baby where there is no apparent cause of death. Studies have shown that sleeping on the stomach increases a baby's risk of SIDS. *Is there an increased risk of a baby dying from SIDS if the mother smokes during pregnancy?* (SIDS)	Yes
5. Breast cancer is one of the most common types of cancer found in women. *Is there an increased chance of developing breast cancer for women who have a family history of breast cancer?* (Breast cancer)	Yes
6. Men are encouraged by our culture to be tough. Unfortunately, many men tend to think that asking for help is a sign of weakness. *In Australia, do more men die by committing suicide than women*? (Suicide)	Yes
7. Many people use home therapies when they are sick or to keep healthy. Examples of home therapies include drinking chicken soup when sick, drinking milk before bed for a better night's sleep, and taking vitamin C to prevent the common cold. *Is there evidence to support the taking of vitamin C supplements to help prevent the common cold?* (Cold)	No
8. We know that we can catch AIDS from bodily fluids, such as from needle sharing, having unprotected sex, and breast-feeding. We also know that some diseases can be transmitted by mosquito bites. *Is it likely that we can get AIDS from a mosquito bite?* (AIDS)	No

Data Analysis

Subjects' searches and their selected documents, pre-/post-search answers and confidence, post-feedback responses, time taken from answering the question pre-search to answering post-search, and responses to the pre-search and post-search questionnaire were logged during the experiment. Responses to questions were coded as "correct," "don't know," or "incorrect" according to the predetermined answers for each question. All cases in which subjects did not conduct a search before providing an answer or seeking the social feedback, did not answer the question post-search, or answered "don't know" post-search were removed from the data analysis.

The test for difference between proportions was used to compare differences between subjects' pre-search, post-search, and post-feedback answers and to compare changes in confidence in answers pre- and post-search. The chi-square test was used to examine whether there was a statistically significant relationship between subjects' confidence in their post-search answers and their tendency to change answers after feedback with other subjects' answers. The McNemar test was used to examine the direction of change in pre- and post-feedback answers.

Results

Subjects and Sample

After data exclusion (Figure 2), the study consisted of 211 subjects who made 928 responses, 1606 searches, and 3019 document accesses. Table 2 presents demographic attributes and self-rated search skills and frequency of searching the Web for general topics and health-related issues. Overall, subjects on average took 361 seconds (SD 281.2) to search, made 1.73 (SD 1.391) searches, and accessed 3.25 (SD 3.067) documents to answer a question.

Figure 2. Data exclusion procedure

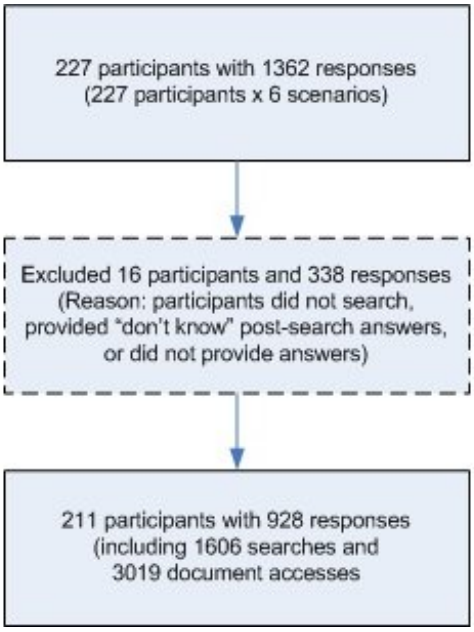


Table 2. Characteristics of subjects (N = 211)

Characteristic	No. (%)
Gender	
Female	130 (61.6)
Male	81 (38.4)
Age (years)	
<25	139 (65.9)
25 to 34	46 (21.8)
35 to 44	12 (5.7)
≥ 45	14 (6.6)
Search skill	
Fair or poor	46 (21.8)
Good	100 (47.4)
Very good	65 (30.8)
Search frequency	
Once a week or less	13 (6.2)
Several times a week	198 (93.8)
Health search frequency	
Never	9 (4.3)
Less than once a week	94 (44.5)
Once a week	52 (24.6)
Several times a week	56 (26.5)

Impact on Decision Accuracy

As shown in Table 3, most subjects, 56.5% (95% CI: 53.3-59.6), answered correctly both before and after searching, which was termed right-right (RR). This was followed by 25.5% (95% CI: 22.8-28.4) who improved their answers after searching, wrong-right (WR), 13.3% (95% CI: 11.2-15.6) who never answered correctly, wrong-wrong (WW), and 4.7% (95% CI: 3.6-6.3) who went from right to wrong (RW).

The test for difference between proportions shows that there was a statistically significant improvement (21%) in the percentage of correct answers before and after searching(pre-search 61.2% [95% CI: 58.0-64.3]; post-search

82.0% [95% CI: 79.4-84.3]; z= −1.21, P <.001). There was also a marginal significant improvement in the percentage of correct answers before and after feedback with other subjects' answers (pre-feedback 82.0% [95% CI: 79.4-84.3]; post-feedback 85.3% [95% CI: 82.9-87.5]; z= −1.95, P =.051; Table 4).

Table 3. Changes in answer before and after searching (N = 928; adapted from [25])

Pre- Search	Post-Search	Percentage (95% CI)	Total No.
Right	Right	56.5 (53.3-59.6)	524
Wrong	Right	25.5 (22.8-28.4)	237
Wrong	Wrong	13.3 (11.2-15.6)	123
Right	Wrong	4.7 (3.6-6.3)	44

Table 4. Correct answers by case scenario (N = 928)

Case Scenario (n)	Correct Before Searching, No. (%)	Correct After Searching, No. (%)	Correct After Feedback, No. (%)
Diet (115)	38 (33.0)	72 (62.6)	79 (68.7)
Hepatitis B (123)	90 (73.2)	108 (87.8)	114 (92.7)
Alcohol (113)	93 (82.3)	94 (83.2)	99 (87.6)
SIDS (111)	71 (64.0)	95 (85.6)	97 (87.4)
Breast cancer (121)	108 (89.3)	108 (89.3)	111 (91.7)
Suicide (113)	63 (55.8)	98 (86.7)	104 (92.0)
Cold (111)	22 (19.8)	68 (61.3)	71 (64.0)
AIDS (121)	83 (68.6)	118 (97.5)	117 (96.7)
Total (928)	568 (61.2)	761 (82.0)	792 (85.3)

Impact of Confidence

Table 5 shows that the most frequently self-reported change in confidence for all responses before and after searching was "increased confidence" (WW 51.9% [95% CI: 42.5-61.0], WR 54.0% [95% CI: 46.3-61.6], RW 40.4% [95% CI: 27.6-54.7], RR 71.1% [95% CI: 67.4-74.6]).

More than half of subjects (55.6%; 95% CI: 37.3-72.4) who did not know the answer pre-search and answered incorrectly post-search (DW) reported that they were confident or very confident with their incorrect post-search answer (Table 6). In fact, 82.0% (95% CI: 75.5-87.1) of subjects who were incorrect post-search reported being confident or very confident with their post-search answer (Table 7). Although Table 7 shows that the proportion of subjects with highly confident correct answers (ie, confident or very confident) significantly increased after searching (pre-search 61.6% [95% CI: 57.6-65.5]; post-search 95.5% [95% CI: 93.8-96.8]; z= –15.60, P <.001), the proportion of subjects with highly confident incorrect answers also increased after searching (pre-search 55.3% [95% CI: 50.1-60.3]; post-search 82.0% [95% CI: 75.5-87.1]; z= –6.75, P <.001).

Table 5. Changes in confidence in original answer following searches (N = 905; adapted from [26])*

Change in Confidence	WW † (n = 108), No. (%)	WR † (n = 161), No. (%)	RW (n = 47), No. (%)	RR (n = 589), No. (%)
Decreased	15 (13.9)	58 (36.0)	14 (29.8)	5 (0.8)
No change	37 (34.3)	16 (9.9)	14 (29.8)	165 (28.0)
Increased	56 (51.9)	87 (54.0)	19 (40.4)	419 (71.1)

*In 23 responses, subjects did not report a confidence rating.

†Includes subjects who did not know the answer before searching.

Table 6. Confidence in post-search answer for subjects who did not know answer before searching (N = 147; adapted from [26])

Post-Search Confidence	Wrong After Search (n = 27), No. (%)	Right After Search (n = 120), No. (%)
Not confident /somewhat confident	12 (44.4)	13 (1.8)
Confident /very confident	15 (55.6)	107 (89.2)

Table 7. Comparison of confidence between pre-search and post-search right and wrong answers (N = 928)

Confidence in Answer	Pre-Search, No. (%)	Post-Search, No. (%)	*z* Score	*P* Value
Right answer	(n= 568)	(n = 761)		
Not confident/ somewhat confident	208 (36.6)	34 (4.5)	14.91	< .001
Confident/ very confident	350 (61.6)	727 (95.5)	–15.60	< .001
Not provided	10 (1.8)	–	–	–
Wrong answer	(n= 360)	(n = 167)		
Not confident/ somewhat confident	154 (42.8)	30 (18.0)	6.28	< .001
Confident/ very confident	199 (55.3)	137 (82.0)	–6.75	< .001
Not provided	7 (1.9)	–	–	–

Impact of Social Feedback

Those who were not as confident in their post-search answers were 28.5% more likely than those who had higher levels of confidence to change their answer after feedback with other subjects' post-search answers (not confident / somewhat confident 34.4% [95% CI: 23.9-46.6]; confident / very confident 5.9% [95% CI: 4.5-7.7]; χ^2_1= 66.65, *P* <.001; Table 8). Those who changed their answer after feedback were more likely to change it from wrong to right than from right to wrong (McNemar χ^2_1= 15.25, *P* <.001; Table 9).

Table 8. Number of subjects who changed their post-search answer after feedback (N = 928)

Post- Search Confidence	Changed Answer , No. (%)	Did Not Change Answer , No. (%)
Not confident/ somewhat confident (n = 64)	22 (34.4)	42 (65.6)
Confident/ very confident (n = 864)	51 (5.9)	813 (94.1)

Table 9. Changes in post-search answer before and after feedback (N = 928)

Before Feedback	After Feedback	
	Right, No. (%)	Wrong, No. (%)
Right (n= 167)	122 (73.1)	45 (26.9)
Wrong (n= 761)	14 (1.8)	747 (98.2)

Discussion

This research demonstrates that while health care consumers can improve the accuracy of their answers to health care questions after searching quality online resources, their confidence in answers is not a good indicator of the answer being correct. Further, consumers who are not confident in their answers after searching are more likely to be influenced to change their views after feedback with other consumers' answers.

Results of this study for nonclinically trained users are in line with studies that reported search engines can improve the ability of clinically trained users to answer questions [25,27,28]. The 21% improvement in accuracy between pre-search and post-search answers reported in this study corresponds with the study conducted by Hersh et al [28], which found that 66 medical and nurse practitioner students were able to improve their answers to a set of five clinical questions by up to 20% before and after using Medline. Our improvement rate also corresponds with the 21% improvement reported for clinicians who used the same search engine to answer eight clinical scenario questions in a controlled laboratory setting (pre-search correct 29% [95% CI: 25-33]; post-search correct 50% [95% CI: 46-54]; *z*= 9.58, *P*< .001) [25].

Findings from this research and previous studies have shown that confidence is not always a good indicator of decision accuracy [26,29]. The observation that 55.6% (95% CI: 37.3-72.4) of subjects in this study who did not know the answer before searching reported being confident or very confident in their incorrect post-search answers (DW) concurs with the result reported by Westbrook et al [26], which found that among clinicians who did not know the answer before searching and were incorrect after searching (DW), 60% of doctors and 52% of clinical nurse consultants reported being confident or very confident in their incorrect post-search answer. This has implications for large-scale national surveys (such as those conducted by the Pew Research Center), which often use confidence as a metric to infer public opinion. In addition, confidence often shapes the way people make decisions (eg, in the form of the overconfidence bias [30,31]), and studies have shown that people can experience cognitive biases while searching for online information to answer questions [32]. These biases, such as the anchoring and order effects, can influence the way people attend to and process information to make a

decision. More research is needed to help users assess the impact of their levels of confidence and understand how their confidence might be shaping their beliefs and ability to attend to new information.

Our findings on the impact of social feedback also concur with studies that report people are one of the important sources of information that influence clinicians' and health care consumers' actions when confronted with a clinical or health-related concern [19,20,33-36]. With the role of the Internet as a social network, typified by the growing interest in sites like Wikipedia, FaceBook, and MySpace, we can envisage more consumers seeking health-related information and advice from online peer networks. However, there appears to be no prior study that has evaluated the health care impact of the social feedback that is possible through such websites. In addition, it is now clear that it is not sufficient to just provide access to reliable online resources for health care consumers. The decisions consumers make are shaped by their confidence and by the influence of their peers and broader social community. Our research suggests that connecting consumers to trustworthy and relevant networks of human resources could be a significant addition to online health resources. As consumers play an increasingly active role in managing their health, it is important not to underestimate the extent to which online search engines and online peer networks can influence the way people manage their health care.

Acknowledgments

The authors would like to thank Dr. David Thomas and Dr. Ilse Blignault for their assistance in subject recruitment and the development of the health care consumer case scenarios. This research was supported by the Australian Research Council SPIRT grant and APAI scholarship C00107730. The Quick Clinical search engine used in the study was developed with support from NHMRC project grant 300435 and the NHMRC development grant 300591. None of the funding sources had any 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.

Conflicts of Interest

The University of New South Wales and Enrico Coiera could benefit from the commercial exploitation of the Quick Clinical search engine or its technologies.

References

1. Hersh WR. Evidence-based medicine and the Internet. ACP J Club 1996;5(4):A14-A16.
2. Eysenbach G, Jadad AR. Evidence-based patient choice and consumer health informatics in the Internet age. J Med Internet Res 2001 Jun 12;3(2):E19 [FREE Full text] [Medline: 11720961] [doi: 10.2196/jmir.3.2.e19]
3. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700 [FREE Full text] [Medline: 12020305] [doi: 10.1001/jama.287.20.2691]
4. McCray AT. Promoting health literacy. J Am Med Inform Assoc 2005;12(2):152-163 [FREE Full text] [Medline: 15561782] [doi: 10.1197/jamia.M1687]
5. Slater MD, Zimmerman DE. Descriptions of Web sites in search listings: a potential obstacle to informed choice of health information. Am J Public Health 2003 Aug;93(8):1281-1282 [FREE Full text] [Medline: 12893614]
6. Ilic D, Bessell TL, Silagy CA, Green S. Specialized medical search-engines are no better than general search-engines in sourcing consumer information about androgen deficiency. Hum Reprod 2003 Mar;18(3):557-561 [FREE Full text] [Medline: 12615824] [doi: 10.1093/humrep/deg154]
7. Keselman A, Tse T, Crowell J, Browne A, Ngo L, Zeng Q. Assessing consumer health vocabulary familiarity: an exploratory study. J Med Internet Res 2007;9(1):e5 [FREE Full text] [Medline: 17478414] [doi: 10.2196/jmir.9.1.e5]
8. Charnock D, Shepperd S. Learning to DISCERN online: applying an appraisal tool to health websites in a workshop setting. Health Educ Res 2004 Aug;19(4):440-446 [FREE Full text] [Medline: 15155597] [doi: 10.1093/her/cyg046]
9. Zeng QT, Crowell J, Plovnick RM, Kim E, Ngo L, Dibble E. Assisting consumer health information retrieval with query recommendations. J Am Med Inform Assoc 2006;13(1):80-90 [FREE Full text] [Medline: 16221944] [doi: 10.1197/jamia.M1820]
10. Marshall L, Williams D. Health information: does quality count for the consumer? How consumers evaluate the quality of health information materials across a variety of media. J Libr Inf Sci 2006;38(3):141-156.
11. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371 [FREE Full text] [Medline: 15224975]
12. Schwartz KL, Roe T, Northrup J, Meza J, Seifeldin R, Neale AV. Family medicine patients' use of the Internet for health information: a MetroNet study. J Am Board Fam Med 2006;19(1):39-45 [FREE Full text] [Medline: 16492004]
13. Peterson G, Aslani P, Williams KA. How do consumers search for and appraise information on medicines on the Internet? A qualitative study using focus groups. J Med Internet Res 2003;5(4):e33. [Medline: 17478414] [doi: 10.2196/jmir.5.4.e33]

14. Dutta-Bergman M. Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. J Med Internet Res 2003 Sep 25;5(3):e21 [FREE Full text] [Medline: 14517112] [doi: 10.2196/jmir.5.3.e21]
15. Anderson JG. Consumers of e-health: patterns of use and barriers. Soc Sci Comput Rev 2004;22(2):242-248. [doi: 10.1177/0894439303262671]
16. Kivits J. Informed patients and the internet: a mediated context for consultations with health professionals. J Health Psychol 2006 Mar;11(2):269-282. [Medline: 16464924] [doi: 10.1177/1359105306061186]
17. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734 [FREE Full text] [Medline: 12885689] [doi: 10.1001/archinte.163.14.1727]
18. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624 [FREE Full text] [Medline: 16344419] [doi: 10.1001/archinte.165.22.2618]
19. 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 [FREE Full text] [Medline: 11926853] [doi: 10.1001/archinte.162.7.792]
20. Lieberman MA, Golant M, Giese-Davis J, Winzlenberg A, Benjamin H, Humphreys K, et al. Electronic support groups for breast carcinoma: a clinical trial of effectiveness. Cancer 2003 Feb 15;97(4):920-925 [FREE Full text] [Medline: 12569591] [doi: 10.1002/cncr.11145]
21. Coiera E, Walther M, Nguyen K, Lovell NH. Architecture for knowledge-based and federated search of online clinical evidence. J Med Internet Res 2005;7(5):e52 [FREE Full text] [Medline: 16403716] [doi: 10.2196/jmir.7.5.e52]
22. PubMed. URL: http://www.pubmed.gov/ [accessed 2008 Jan 9] [WebCite Cache ID 5Tger8qKj]
23. MedlinePlus. URL: http://medlineplus.gov/ [accessed 2008 Jan 9] [WebCite Cache ID 5TgfWSJD5]
24. HealthInsite. URL: http://www.healthinsite.gov.au/ [accessed 2008 Jan 9] [WebCite Cache ID 5TgfggXeD]
25. Westbrook JI, Coiera EW, Gosling AS. Do online information retrieval systems help experienced clinicians answer clinical questions? J Am Med Inform Assoc 2005;12(3):315-321 [FREE Full text] [Medline: 15684126] [doi: 10.1197/jamia.M1717]
26. Westbrook JI, Gosling AS, Coiera EW. The impact of an online evidence system on confidence in decision making in a controlled setting. Med Decis Making 2005;25(2):178-185. [Medline: 15800302] [doi: 10.1177/0272989X05275155]
27. Hersh WR, Crabtree MK, Hickam DH, Sacherek L, Rose L, Friedman CP. Factors associated with successful answering of clinical questions using an information retrieval system. Bull Med Libr Assoc 2000 Oct;88(4):323-331 [FREE Full text] [Medline: 11055299]
28. Hersh WR, Crabtree MK, Hickam DH, Sacherek L, Friedman CP, Tidmarsh P, et al. Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. J Am Med Inform Assoc 2002;9(3):283-293 [FREE Full text] [Medline: 11971889] [doi: 10.1197/jamia.M0996]
29. Pallier G, Wilkinson R, Danthiir V, Kleitman S, Knezevic G, Stankov L, et al. The role of individual differences in the accuracy of confidence judgments. J Gen Psychol 2002 Jul;129(3):257-299. [Medline: 12224810]
30. Kahneman D, Slovic P, Tversky A. Judgment under Uncertainty: Heuristics and Biases. New York: Cambridge University Press; 1982.
31. Wickens CD, Hollands JG. Decision Making. Engineering Psychology and Human Performance. 3rd edition. Upper Saddle River, NJ: Prentice Hall; 2000:293-336.
32. Lau AYS, Coiera EW. Do people experience cognitive biases while searching for information? J Am Med Inform Assoc 2007;14(5):599-608. [Medline: 17600097] [doi: 10.1197/jamia.M2411]
33. Dawes M, Sampson U. Knowledge management in clinical practice: a systematic review of information seeking behavior in physicians. Int J Med Inf 2003;71(1):9-15. [doi: 10.1016/S1386-5056(03)00023-6]
34. Coumou HCH, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. J Med Libr Assoc 2006 Jan;94(1):55-60 [FREE Full text] [Medline: 16404470]
35. Shaw BR, McTavish F, Hawkins R, Gustafson DH, Pingree S. Experiences of women with breast cancer: exchanging social support over the CHESS computer network. J Health Commun 2000;5(2):135-159. [Medline: 11010346] [doi: 10.1080/108107300406866]
36. McGettigan P, Golden J, Fryer J, Chan R, Feely J. Prescribers prefer people: The sources of information used by doctors for prescribing suggest that the medium is more important than the message. Br J Clin Pharmacol 2001 Feb;51(2):184-189 [FREE Full text] [Medline: 11259994] [doi: 10.1111/j.1365-2125.2001.01332.x]

Abbreviations

DW: didn't know, wrong
RR: right, right
RW: right, wrong

UNSW: University of New South Wales
WR: wrong, right
WW: wrong, wrong

submitted 01.08.07; peer-reviewed by D Hansen; comments to author 06.11.07; revised version received 30.11.07; accepted 04.01.08; published 22.01.08.

Please cite as:
Lau AYS, Coiera EW
Impact of Web Searching and Social Feedback on Consumer Decision Making: A Prospective Online Experiment
J Med Internet Res 2008;10(1):e2
URL: http://www.jmir.org/2008/1/e2/
doi:10.2196/jmir.963
PMID:18244893

Original Paper

What Do Evaluation Instruments Tell Us About the Quality of Complementary Medicine Information on the Internet?

Matthew Breckons, BSc; Ray Jones, PhD; Jenny Morris, PhD; Janet Richardson, PhD

[1]School of Nursing and Community Studies, University of Plymouth, Plymouth, UK

Corresponding Author:
Ray Jones, PhD
School of Nursing and Community Studies
University of Plymouth
Faculty of Health and Social Work
Plymouth PL4 8AA
United Kingdom
Phone: +00 44 1752 233886
Fax: +00 44 1752 586748
Email: ray.jones@plymouth.ac.uk

Abstract

Background: Developers of health information websites aimed at consumers need methods to assess whether their website is of "high quality." Due to the nature of complementary medicine, website information is diverse and may be of poor quality. Various methods have been used to assess the quality of websites, the two main approaches being (1) to compare the content against some gold standard, and (2) to rate various aspects of the site using an assessment tool.

Objective: We aimed to review available evaluation instruments to assess their performance when used by a researcher to evaluate websites containing information on complementary medicine and breast cancer. In particular, we wanted to see if instruments used the same criteria, agreed on the ranking of websites, were easy to use by a researcher, and if use of a single tool was sufficient to assess website quality.

Methods: Bibliographic databases, search engines, and citation searches were used to identify evaluation instruments. Instruments were included that enabled users with no subject knowledge to make an objective assessment of a website containing health information. The elements of each instrument were compared to nine main criteria defined by a previous study. Google was used to search for complementary medicine and breast cancer sites. The first six results and a purposive six from different origins (charities, sponsored, commercial) were chosen. Each website was assessed using each tool, and the percentage of criteria successfully met was recorded. The ranking of the websites by each tool was compared. The use of the instruments by others was estimated by citation analysis and Google searching.

Results: A total of 39 instruments were identified, 12 of which met the inclusion criteria; the instruments contained between 4 and 43 questions. When applied to 12 websites, there was agreement of the rank order of the sites with 10 of the instruments. Instruments varied in the range of criteria they assessed and in their ease of use.

Conclusions: Comparing the content of websites against a gold standard is time consuming and only feasible for very specific advice. Evaluation instruments offer gateway providers a method to assess websites. The checklist approach has face validity when results are compared to the actual content of "good" and "bad" websites. Although instruments differed in the range of items assessed, there was fair agreement between most available instruments. Some were easier to use than others, but these were not necessarily the instruments most widely used to date. Combining some of the better features of instruments to provide fewer, easy-to-use methods would be beneficial to gateway providers.

(J Med Internet Res 2008;10(1):e3) doi:10.2196/jmir.961

KEYWORDS

Consumer Health Informatics; Internet; quality of information; complementary medicine

Introduction

While the ever-expanding source of health information might be seen as a positive step in consumer empowerment (between 36% and 55% of Internet users access online health information [1-4]), several studies have highlighted problems with the quality of the information [5-8]. Searching for relevant and reputable complementary medicine information is particularly challenging [9], in part due to methodological challenges. Schmidt and Ernst [10] report that claims are made that can put consumers at risk; in some cases, adherence to advice obtained from the Internet has had serious consequences [8,11]. This is of particular concern for people who may be vulnerable, such as those affected by cancer [12]. In the case of hydrazine sulfate poisoning [8], inaccurate and exaggerated claims of effectiveness and lack of information on side effects were blamed for misleading a consumer who assumed the substance was safe.

A 2002 study estimated that 5 million adults in England lack basic literacy [13]. Furthermore, understanding health information may be a complex process requiring more than basic literacy skills as even well-educated people can have difficulties making sense of it [14]. A US study of over 350 health sciences students showed that while many rated themselves as possessing good research skills, only a small proportion were able to demonstrate that they could identify reliable information [15]. While consumers may regularly make judgments of the quality of information received through traditional media such as newspapers, books, or leaflets, quality indicators for Internet content may not be as evident to users [16].

Consumers looking for health information are likely to select the first few links that appear on search engines and tend not to look for information about site authors or disclaimers that sites may make [17]. Studies have found that when consumers evaluate the quality of health information on the Internet, they tend to rely onendorsement by government agencies or professional organizations, their own perception of reliability of the website source, and the understandability of the information [18,19].

Several strategies have been designed to help health information seekers access high-quality information, including codes of conduct, gateway sites (portals), and evaluation instruments. The development of instruments has received the greatest attention, and some suggest that their use by consumers can educate the user as to the characteristics of a good quality website [16], but the behavior of the majority of consumers would suggest that gateways may be the best approach. Evaluation instruments can still provide a method for researchers to help choose links for gateways. Evaluation instruments work on the premise that they can identify “quality” sites on the assumption that sites that conform to indicators of quality are likely to contain accurate information. Accurate information is defined as being based on a gold standard of information in the field. While it is not possible for someone with no domain knowledge to assess accuracy, it may be that instruments can be used to help make judgments on quality and hence predict accuracy.

Gateways are collections of sites that have been prescreened and deemed of high enough quality to be approved by a governing organization. Examples of these are Healthfinder [20] and Intute [21]. Although maintaining portals can be labor intensive, organizations providing services such as complementary medicine for cancer need to be able to recommend sites to their patients. There are many instruments that have been designed, ranging from simple checklists to long and complex documents providing detailed accounts of assessment methodologies, and organizations running a portal need to choose which to use. Three characteristics that would seem important are (1) agreement with other instruments when rating a website, (2) ease of use, and (3) longevity. On the latter, many instruments seem to have a very limited life span. In 1998, for example, Jadad and Gagliardi identified 47 instruments used to rate the quality of health information on the Internet [22], but 4 years later [23], only six of these instruments still existed.

Our study was conducted in a center providing complementary care for people affected by cancer. The aim of this study was to identify website evaluation instruments and to assess their performance when used by a researcher to evaluate a sample of 12 websites on complementary medicine for people with breast cancer. In particular, we asked the following:

- Do the instruments use the same criteria, and do they agree on the ranking of websites?
- How easy are the different instruments for a researcher to use?
- Are these instruments likely to remain in use such that future readers will appreciate the assessment method used?
- Could we identify a pragmatic approach to identify good quality complementary medicine websites using existing instruments?

Methods

Literature Search for Evaluation Instruments

We defined an evaluation instrument as something that an Internet user could use to assess the quality of a website containing health information. To identify evaluation instruments, search terms were based on previous papers that had attempted to identify instruments for evaluating the quality of Internet health information [22,24,25]. The databases Medline, AMED, BNI, CINAHL, EMBASE, and PsychInfo were searched in February 2007 using the following terms: “evaluat* OR assess* OR rating OR rat* OR ranking OR rank* OR quality OR criteria AND website* OR world wide web OR Internet.” This achieved results of 29,622, 233, 123, 14,859, 8678, and 10,593, respectively. When in excess of 1000, the most recent results from each database were examined.

In addition to the above databases, the search engines Google, MSN, Yahoo, and WebCrawler were searched using the following terms: “evaluate OR assess OR rating OR criteria OR quality AND websites OR Internet.” This achieved results of 212,000,000, 25,410,704, 38,400,000, and 28, respectively. With the exception of WebCrawler (28 results), the first 100

results of each were examined. Relevant research papers and bibliographies were also examined for relevant references. Several large studies that had attempted to systematically search for and identify instruments for assessing Internet health information were found [22-25].

Instruments were selected if they provided the user with explicit instructions for evaluating the quality of a website containing health information. While the HON Code of Conduct (HONcode) has been mentioned as a gateway site, its evaluation criteria were included in this report.

Internet Search for Complementary Medicine Websites

To search for websites to be assessed, a search for "complementary (medicine OR therapies) AND breast cancer" was performed in February 2007 using the Google search engine. This resulted in 1,170,000 hits. The first six results were selected on the basis that people are most likely to look at only the first few results produced by a search engine [17]. Another six results were chosen purposively to obtain a selection of sites with different purposes and origins: sites belonging to charities, sponsored sites, and sites selling products.

Assessment of Websites

The 12 websites were evaluated using each of the 12 evaluation instruments (ie, 144 assessments). Each site was given a mark using the individual scoring system for each instrument, which was then converted to a percentage score. Some instruments gave negative scores for failing to meet criteria; therefore, it was possible for a negative score to be obtained. Sites were then ranked from 1 (best) to 12 (worst) based on these scores.

Comparison of Evaluation Instruments

The range of criteria used by the identified instruments was compared to the nine main criteria identified by the Health Improvement Institute and Consumer Reports WebWatch (HIICRW) [26] in 22 health information rating instruments. Agreement between instruments was assessed by a correlation matrix using Spearman rank correlation on the instruments' ranking of the websites.

Illustrative Comparison of Best and Worst Sites

The range of content on each site made comparison against a gold standard impossible. Nevertheless, we sought some "face validity" in that sites ranked as "good" or "poor" using these evaluation instruments matched with common sense. Statements made on the site ranked the best by the sum of the 12 instruments were compared to those on the site ranked the worst.

Citation Search for Use of Evaluation Instruments

A citation search on Web of Science was carried out using the original papers describing the instruments. A sample of papers that cited the original paper was reviewed, and an estimate was made of the number of papers that had used the tool. A citation search on Google using the instrument's http address was carried out. A sample of websites that cited the original Web address of the tool was checked to see if the citation was correct. The number of citations on Web of Science or Google was classified as low (less than 10), medium (11-100), or high (greater than 100).

Longevity of Instruments

The URLs of instruments that had been identified in four previous studies were checked (as part of the literature search) to see if they still existed. Instruments were reported as unavailable if the original URL was not found and searching the original site or Google for the instrument did not locate it.

Results

Evaluation Instruments Available

A total of 39 instruments that disclosed their criteria and aimed to help users identify good quality information online were identified. Of these, 12 met our inclusion criteria (Table 1); the other 27 were excluded (Table 2). Instruments were selected if they provided the user with a set of objectives and closed questions that could be applied to a website containing health information by someone with no prior subject knowledge and without having to look at sources other than the website being assessed. Reasons for exclusion of the 27 instruments included the following:

- A consumer could not apply the instrument without further knowledge (eg, "Is the information written by reputable authors?").
- Scoring details were unavailable (eg, Instructions stated to score each criterion on a scale of 1-5, but no further information was given as to how to allocate a value.).
- Questions were not objective (eg, "Are the graphics attractive?").
- Instrument was not designed specifically for health information.
- Questions were open ended (eg, "What are the author's qualifications?").
- Instrument took the role of a tutorial that gave tips on how to find reliable health information on the Internet but was not applicable as an instrument.

Websites Sampled

Table 3 shows the websites that were rated using the instruments; four of the sites were run by UK charities, two sites were selling products, and three sites were US sites offering cancer treatment. One site was run by a network of health professionals, one site was funded by advertising on its site, and one site was funded by sponsors.

Table 1. Evaluation instruments used in the study

Evaluation Instrument	Method of Assessment	Ease of Use (Researcher Assessment)	Comprehensiveness (HIICRW Criteria Met, out of 9)
WEB FEET HEALTH Collection: Criteria for Site Selection (WEB FEET) [27]	24 statements: agree or disagree	+ Straightforward questions − Time consuming to apply	7
HONcode [28]	8 desirable properties	+ Short tool, quick to apply + Each element includes guidelines	4
Emory University Rollins School of Public Health, Health-Related Web Site Evaluation Form (Emory) [29]	36 statements: +1 disagree, +2 agree, 0 N/A Score: 0-60	+ Interpretation of score − Time consuming	7
University of Michigan Web Site Evaluation Checklist (Michigan) [30]	43 questions, with variety of positive and negative scores Score: −80 to +80	+ Printable rating form + Interpretation of score − Time consuming − Complex scoring system	6
Kellogg Library (University of Dalhousie), Evaluation of Health Information on the Internet (Kellogg) [31]	31 questions: agree or disagree	+ Simple questions, straightforward to use − Time consuming	8
DISCERN Quality Criteria for Consumer Health Information (DISCERN) [32]	16 questions on 5-point analogue scale from “No” to “Yes” Overall score: 1-5	+ Explanation of criteria + Interpretation of score − Answers on visual analogue scale more difficult / time consuming	5
National Center for Complementary and Alternative Medicine (NCCAM), 10 Things to Know About Evaluating Medical Resources on the Web (NCCAM) [33]	10 questions each with explanation	+ Clear guidance of how to use criteria − No scoring system	6
US Pharmacist tool (Pharm) [34]	15 questions with yes/no answers	+ Easy to apply − No scoring system	6
Minervation Validation Instrument for Health Care Web Sites (Minervation) [35]	Semi-automated tool requires URL of site being assessed Drop-down menus to answer questions of content and usability Rating automatically calculated Score: 0-100%	+ Automated usability check + Drop-down menus, fast + Interpretation of score	5
Nicoll LH, author’s guidelines (Nicoll) [36]	Mnemonic (PLEASED) with yes/no questions, each with author justification of importance	+ Quick to use + Explanation of criteria − No scoring system	5
Silberg et al, authors’ guidelines (Silberg) [37]	4 items that should be met	+ Quick to apply − Does not assess aspects unique to Internet information	3
Sandvik score (Sandvik) [38]	7 questions, each with 3 options scored 0-2 Score: 0-14	+ Quick to apply + Simple scoring system	3

Table 2. Evaluation instruments excluded

Evaluation Instrument	Requires Further Knowledge	Scoring Details Unavailable	Questions not Objective	Not Health Specific	Open-Ended Questions	Tutorial
Quality Criteria for Health Related Websites [39]	x					
Net Scoring Criteria to Assess the Quality of Health Internet Information [40]		x				
Criteria for Evaluating the Quality of Health Information on the Internet [41]	x		x			
Administration Design Quality Web Site Evaluation Method [42]	x					
Evaluating Websites [43]			x	x		
Navigating the Health Care System: How to Evaluate Health Information on the Internet [44]					x	x
Rating Criteria and Excellence Awards [45]			x			
Clean Bill of Health Award [46]				x		
Health Website Rating (HWR) Project: HII Health Website Rating Instrument (HWRI) [47]					x	
Clearing House*		x				
Best of the Web in Mental Health: Rating Guidelines [48]			x			
Commentary: Measuring Quality and Impact of the World Wide Web [49]						x
Evaluating Internet Health Information: A Tutorial From the National Library of Medicine [50]						x
MedlinePlus Guide to Healthy Web Surfing [51]						x
Taking Charge of Health Information [52]					x	
How to Evaluate Health Information on the Internet: Questions and Answers [53]					x	
How to Find the Most Trustworthy Health Information on the Internet [54]			x			
Internet Detective [55]			x			x
Internet for Health and Well-Being [56]			x			x
Suggestions for Using the Internet to Find New Cancer Treatments [57]						x
Internet Health Coalition*						x
How to Judge the Quality of a Web Site [58]			x			x
Intute: Health and Life Sciences Evaluation Guidelines [59]			x			x
Best Practice Web Assessments: Evaluation Criteria [60]		x		x		
Evaluation Form Used for LASIK Websites [61]			x			
Quality Standards for Medical Publishing on the Web [62]				x		
Evaluating Internet Resources in Complementary and Alternative Medicine [63]				x		x

*Instruments became unavailable between initial search (February 2007) and final submission of paper (November 2007).

Table 3. Twelve websites on complementary medicine and breast cancer

Website	Purpose of Site	Reason for Inclusion
Breast Cancer Care [64]	A UK charity aimed at providing information and support for people affected by breast cancer. National Health Service (NHS) information partner.	1st on Google
Breast Cancer Haven [65]	A UK charity that runs day centers offering support, information, and complementary therapies to people affected by breast cancer.	2nd on Google
CancerHelp UK [66]	A UK information service for people with cancer and their families run by the Cancer Research UK charity for cancer and cancer care.	3rd on Google
Imaginis [67]	An independent resource for information and news on breast cancer and related women's health topics.	4th on Google
MD Anderson Cancer Center [68]	An information service run by the University of Texas MD Anderson Cancer Center that offers medical services to people with cancer.	5th on Google
Cancer Treatment Centers of America [69]	An information service run by Cancer Treatment Centers of America, a network of cancer treatment hospitals and facilities offering conventional and complementary therapies.	6th on Google
Cancerbackup [70]	A cancer information charity offering information, practical advice, and support for cancer patients, their families, and caregivers.	Charity
Heart Spring [71]	A resource for alternative and complementary health information funded by advertising and product sales.	Sponsored: product advertisements
Issels Treatment [72]	Information produced by Issels Medical Center, a private organization offering alternative treatment for cancer.	Private cancer center
Alternative Cancer [73]	A site run by an individual selling a guide to complementary and alternative cancer treatments.	Commercial
MedicineNet [74]	Medical information written by a network of medical professionals.	Sponsored: product advertisements
Elbee Global [75]	A site selling herbal medicines for people with cancer.	Commercial

Assessment of Evaluation Instruments

Comprehensiveness

The HIICRW [26] defined nine criteria that an assessment should have. Assessment of each evaluation instrument against the HIICRW criteria showed considerable variation, implying little consensus on quality markers for websites. Although assessment of more criteria may not mean an evaluation instrument is superior, it is interesting that two of the better-known instruments (HONcode and DISCERN) assessed relatively few of the items described by HIICRW (see Table 1).

Ease of Use

Table 1 shows the researcher's subjective view on the evaluation instruments' ease of use. Time taken is an important component of ease of use; answering Michigan University's 43 questions was extremely time consuming, in contrast to the automated Minervation instrument, which could be applied very quickly. Some instruments were not designed to provide numerical scores. It was useful to have some interpretation of how many criteria a website should meet for it to be thought of as being good or bad quality. Instruments varied in the explanation of their criteria. It was helpful to have further guidance available to answer questions, such as provided by HONcode and DISCERN.

Ranking of Websites

Table 4 shows the percentage score for each of the 12 websites and the ranking from best (1) to worst (12) by each instrument and overall. It was notable that the well-known UK charity site Cancerbackup came only 4th in the overall ranking and that the WEB FEET tool ranked it 7th, way behind the Elbee Global website. The HONcode ranked it 5th, on par with the Elbee Global website. Overall, the best site was Imaginis and the worst, Alternative Cancer.

Comparison of Evaluation Instruments

Table 5 shows the agreement (rank correlations) among instruments on the ranking of the 12 websites from best to worst. Where there is a significant correlation (eg, between Michigan and Kellogg), using either tool would give similar results. This showed that WEB FEET and HONcode seemed to assess different characteristics than the other instruments.

Recognition and Use of Instruments

Recognition, citation, and use of instruments are necessary if they are to survive. Table 6 shows the Web of Science level of citation by other papers describing the instruments and the citations of the instruments' website addresses on Google.

Comparison of Best and Worst Sites

Table 7 shows illustrative extracts of statements made in the best and worst ranked sites. As would be expected, the best site (Imaginis) took a balanced and cautious approach to all claims.

The Alternative Cancer site, rated the worst, made claims that were exaggerated or difficult to prove or disprove.

Longevity of Evaluation Instruments

Table 8 shows four studies that previously searched for and identified evaluation instruments and how many of those instruments were still available in November 2007.

Table 4. Ranking and percentage score of websites

Evaluation Instrument	Breast Cancer Care	Breast Cancer Haven	Cancer Help	Imaginis	MD Anderson	Cancer Treatment Centers	Cancer-backup	Heart Spring	Issels	Alternative Cancer	MedicineNet	Elbee Global
WEB FEET	6	8	1	2	4	10	7	2	10	12	9	4
	75%	67%	92%	83%	79%	58%	71%	83%	58%	46%	63%	79%
HONcode	9	2	9	2	1	5	5	2	9	9	5	5
	50%	88%	50%	88%	100%	63%	63%	88%	50%	50%	63%	63%
Emory	6	7	2	3	3	8	1	8	11	11	3	10
	89%	86%	95%	92%	92%	84%	97%	84%	63%	63%	92%	68%
Michigan	7	6	2	1	2	8	5	9	10	11	4	12
	28%	45%	50%	53%	50%	21%	48%	16%	14%	1%	49%	−14%
Kellogg	5	5	1	1	3	9	4	5	10	11	8	11
	70%	70%	90%	90%	77%	50%	73%	70%	33%	27%	63%	27%
DISCERN	8	4	2	5	2	9	1	5	10	12	7	11
	66%	76%	80%	74%	80%	52%	89%	74%	46%	31%	69%	39%
NCCAM	5	5	3	1	1	10	3	5	5	11	5	12
	60%	60%	70%	90%	90%	50%	70%	60%	60%	40%	60%	20%
Pharm	6	7	2	2	4	9	1	8	12	10	4	10
	80%	73%	93%	93%	87%	60%	100%	67%	33%	47%	87%	47%
Minervation	4	7	2	3	6	5	1	11	9	10	7	12
	73%	62%	79%	74%	71%	64%	82%	46%	60%	48%	62%	34%
Nicoll	3	7	1	3	3	9	1	8	12	9	3	9
	71%	57%	86%	71%	71%	29%	86%	43%	14%	29%	71%	29%
Silberg	7	7	3	3	1	7	3	1	7	7	3	12
	25%	25%	75%	75%	100%	25%	75%	100%	25%	25%	75%	0%
Sandvik	7	7	2	2	1	10	6	2	9	11	2	12
	64%	64%	79%	79%	86%	29%	71%	79%	36%	21%	79%	7%
Overall rank	**8**	**7**	**2**	**1**	**3**	**9**	**4**	**6**	**10**	**12**	**5**	**11**

Table 5. Spearman nonparametric correlation coefficients between evaluation instruments, based on assessment of the websites

	WEB FEET	HONcode	Emory	Michigan	Kellogg	DISCERN	NCCAM	Pharm	Minervation	Nicoll	Silberg	Sandvik
WEB FEET	1.00											
HONcode	.35	1.00										
Emory	.51	.25	1.00									
Michigan	.48	.38	.87*	1.00								
Kellogg	.71*	.39	.84*	.89*	1.00							
DISCERN	.55	.47	.86*	.77*	.87*	1.00						
NCCAM	.55	.39	.78*	.89*	.92*	.82*	1.00					
Pharm	.53	.28	.98*	.88*	.87*	.84*	.80*	1.00				
Minervation	.30	−.02	.85*	.79*	.79*	.70	.75*	.85*	1.00			
Nicoll	.55	.14	.97*	.82*	.83*	.82*	.74*	.97*	.82*	1.00		
Silberg	.51	.51	.61†	.66†	.71†	.71*	.75*	.64†	.37	.59†	1.00	
Sandvik	.59†	.51	.72*	.83*	.82*	.77*	.85*	.73*	.48	.70†	.93*	1.00

*Correlation is significant at the 0.01 level (2-tailed).

†Correlation is significant at the 0.05 level (2-tailed.

Table 6. Number of citations in Web of Science and Google (classified as low, medium and high) suggesting use of instruments (NPI: no paper identified)

Evaluation Instrument	Web of Science	Google
WEB FEET	NPI	Low
HONcode	Med	High
Emory	NPI	Medium
Michigan	NPI	Low
Kellogg	NPI	Low
DISCERN	Med	High
NCCAM	NPI	High
Pharm	NPI	Low
Minervation	NPI	Low
Nicoll	Not cited	Low
Silberg	High	Low
Sandvik	Med	Med

Table 7. Comparison of statements from websites rated best and worst

Best Site (Imaginis)	Worst Site (Alternative Cancer)
"...anecdotal evidence reveals that many alternative or complementary medicines may be beneficial to patients, extensive research is still needed to determine whether non-traditional medicines are truly effective."	"Proven Therapies. Includes a list of successful, long-standing alternative treatments from around the world going unused by the conventional medical system. There is one reason they are the oldest - in the hands of experienced practitioner they work! For example: the very successful nutritional based Gerson therapy. It has been used by untold thousands of people worldwide for over 50 years."
"Chinese herbs have been shown to lessen the side effects of chemotherapyand acupuncture has been shown to reduce nausea (a possible side effect of chemotherapy and other drug therapies)."	"*Every day worldwide, quietly behind the scenes, there are over 100 proven alternative therapies used successfullyagainst cancer. (Get a FREE list of the 78 most popular below) The problem is, nobody bothers to tell the public. Plus, conventional cancer doctors (MD Oncologists) are not taught anything about them in medical schools. This must change!*"
"Not all alternative or complementary medicines are safe."	"The one true secret to success: There are six basic types of proven alternative cancer treatments, and you must use them all together."
"In a recent studypublished in the *Journal of the National Cancer Institute*, researchers found that advanced breast cancerpatients with high stress levels were less likely to live as long as patients who coped well with stress."	"Anvirzel® A new weapon against cancer and AIDS from Ozelle Pharmaceuticals - a herbal extract which is non-toxic and causes no adverse side effects. Closed clinical trials are showing that the drug is especially effective against prostate and breast cancer. The materials of the company promoting Anvirzel. say that Dr Ozel treated 494 cancer patients with the extract, resulting in a high rate of success. The company has organized phase I and II trials in Ireland, and states that the trials confirmed the efficacy of the extract in cancer. They say the patients were improved in their quality of life as well as regression of cancer, reporting no notable side effects. Best results were said to be in prostate, lung and brain cancers. Sarcomas showed stabilization."
"Some preliminary studies have shown that vitamins may help reduce risk of breast cancer or treat the disease."	"Artemisinin A Chinese herb, sweet wormwood (*qinghao* in Chinese). In test tube studies, breast cancer cell research resulted in a 28% reduction of breast cancer cells treated only with artemisinin, and an amazing 98% decrease in breast cancer cells within 16 hours that were treated with artemisinin and an iron-enhancing molecule, transferrin. These treatments had no significant effect on normal human breast cells. This research pointed to the involvement of free iron in the toxic effect of artemisinin toward cancer cells, basically sparing healthy cells. ('Selective toxicity of dihydroartemisinin and holotransferrin toward human breast cancer cells,' Life Sciences 70 {2001) 49-56."

Table 8. Instruments identified in previous studies still available in 2007

Study	Year of Study	No. of Instruments Identified	No. of Instruments Available in November 2007
Jadad and Gagliardi [22]	1998	14	3
Kim et al [25]	1999	27	7
Gagliardi and Jadad [23]	2002	5	1
Bernstam et al [24]	2005	17	3

Discussion

Limitations of This Study

Our study has some limitations. Selection of instruments, website ratings, and HIICRW criteria comparison were performed by only one researcher. Possible interobserver variation may mean that some instruments eligible for inclusion may have been missed and that some excluded may have been included by other reviewers. Due to the nature of the instruments being searched, they do not lend themselves to very specific search terms, meaning that our searches produced many results. Nevertheless, we may have found more tools by examining a greater number of search results or by searching other databases. Two instruments were excluded only for the reason that they were not health specific and, in retrospect, that exclusion criterion may not have been warranted.

Application of the evaluation tools to particular websites may also have produced different results with other researchers. Bernstam et al, in a recent study [76], suggested that some quality criteria may have poor interobserver reliability. However, there is likely to be more variation (both intraobserver and interobserver) in the values attributed to individual characteristics of an assessment tool. When combined to give an overall rank, as we have done in this study, tools are more likely to give consistent results.

What This Study Offers

Although our study has limitations, our experience has a useful message for several groups of people:

- For those assessing or developing gateways who may wish to use an evaluation instrument, this study provides information that may help select an instrument.

- For authors of evaluation instruments, we identified those features that may be desirable to ensure their instrument is useable and useful.
- For information seekers, we show which properties to look for when selecting an instrument and suggest which instruments may be preferable to others.
- For developers of complementary medicine websites, we show the need to use "technical markers of quality" to ensure that their site achieves high scores when assessed by instruments.

Our study also suggests that the popular HONcode may assess quality in a different way than other instruments.

The Quality of Websites

Developers of websites or gateways on complementary medicine need some method to check the quality of what they are presenting, and users of their websites need to be able to assess for themselves, and to believe, the claim that this is a quality website. What does quality mean? Provost et al [77] define quality as the levels of excellence which characterize the content of the site based on accepted standards of quality. At the very least, it should mean that the information presented is evidence based and the evidence is available to be checked.

The Gold Standard Approach

Impicciatorre et al [78] were among the first to assess the reliability of Web page information by comparing it against a gold standard. Others have followed this approach [7,79,80], but in every case, they have been able to focus on specific pieces of information or advice that have an available gold standard. For example, Pandolfini et al [81] compared information on the management of cough in children against a gold standard. Assessing quality in this way is time consuming, and in cases where websites present information on a broader range of topics, not a feasible option. Having some sort of evaluation that allows a quicker test of quality is therefore an attractive option, and for this reason, numerous evaluation instruments have been devised.

Does the Evaluation Instrument Approach Act as Good Proxy for Quality of Information?

Pandolfini et al [81] examined 19 Web pages and noted that no relationship was found between technical aspect, content completeness, and quality of information as compared to a gold standard. However, only one page received a high score on comparison against the gold standard, and this page also scored high on the other two measures. In our study, we have not assessed against a gold standard, but a simple comparison of the content of the best and worst sites using evaluation instruments shows our approach to have face validity. However, we should remain cautious. While instruments are designed to assess the quality of information, they are concerned with quality indicators and can therefore not take into account the accuracy of an individual piece of information. Eysenbach et al [5] are of the opinion that it is unlikely that a universal set of criteria could be developed that would predict the quality of health information websites as there are complex relationships between quality indicators and actual quality of information. While the results of our study suggest that websites rated higher by the evaluation instruments seem typically less likely to contain exaggerated claims, Walji et al [82] analyzed 150 websites dealing with the use of ginseng, ginkgo, and St. John's wort and concluded that domain-independent criteria may not be appropriate for identifying complementary and alternative medicine websites, suggesting that consumers should rely on authoritative providers of information. There may be specific challenges in accessing high-quality information on complementary medicine, but there are several initiatives aimed at providing high-quality, evidence-based information, including the Cancer Specialist Library [83], National Center for Complementary and Alternative Medicine (NCCAM) [84], and Complementary and Alternative Medicine Evidence OnLine (CAMEOL) [85].

Validity, Reliability, and Agreement of Evaluation Instruments

The majority of available instruments have not been tested for reliability [24,86] or validity [86], and few include information describing the development process [82]. DISCERN and the Minervation tool appear to be the only ones that discuss the fact that their instruments have been tested for reliability and validity. Even among researchers, there is likely to be observer variation on various criteria. Bernstam et al [76] examined the degree to which two raters could reliably assess 22 popularly cited quality criteria on a sample of 42 complementary and alternative medicine websites and found poor agreement on 8/22. Good definition of the quality criteria should improve agreement, but the level of agreement between most of the instruments used in this study shows that complete "accuracy" may not be that important. Two of the instruments, HONCode and WEB FEET, did not have good agreement with the other 10 in ranking the best to worst sites. It is not clear why this is. So although HONcode is used frequently, we felt it safer to use those instruments that agreed as most of the other instruments seemed to address most aspects identified by the HIICRW.

Ease of Use

Five of the 12 instruments were time consuming to apply. Bernstam et al [24] took the view that any tool containing more than 10 criteria was too long for routine use and that the majority of available instruments are not user friendly. Although instruments should be comprehensive, and while it may be useful to ask a wide range of questions about a site, it is important that the application of an instrument is practical. Our study suggests that greater coverage of criteria is not necessarily achieved by asking a large number of questions, although if a tool is too short it is unlikely that it could cover a wide range of criteria. There was a great deal of variation in usability of the instruments. The Minervation tool contains an automated feature that allows entry of an URL. It produces an accessibility rating, leaving the user to select answers to questions of reliability and usability from drop-down menus. It then allocates scores for each section, an overall score, and gives a rating of the site in terms of "poor," "fair," or "good." These automated features are in contrast to an instrument such as the one developed by Emory University, which was very time consuming to apply. Some instruments feature further guidance

to assist the user in answering the questions, which was considered a useful attribute.

Range of Criteria

Eight of the instruments contained criteria concerning accessibility; although differing between instruments, this element asked questions about website design, layout, and if there was a search engine included on the page or appropriate links for navigation. While accessibility might not seem directly related to the quality of the information contained in the pages, it is extremely important in terms of the usefulness of the site.

Many websites "lost marks" as they did not display information concerning authorship. Eysenbach et al feel that this may be more related to convention than quality as it is not usual for organizations to display names of individual authors, and this is not necessarily an indicator of quality [5]. The way that the instrument's question is phrased may be crucial in informing users of the quality of a site. Concerning authorship, some would ask "Is the name of the author disclosed?" which, in itself, may show that the site has a good transparency policy, but it does not add clarity to questions of quality as it is still not known if the author is suitably qualified to write on a particular topic. Similarly, regarding currency of the information, "Does the site display the date on which it was last updated?" is not as valuable as "Has the site been updated in the last 6 months?" Hence, an instrument covering the same criteria as another may achieve a different rating due to different wording of its questions.

Number and "Shelf Life" of Evaluation Instruments

Bernstam et al [24] apparently identified 273 instruments; however, they included tools such as "top traffic" that could not be utilized by an Internet user. They identified only seven instruments that could be applied by Internet users. We did not attempt to identify instruments that could not be applied to individual sites by an information seeker.

One problem with any technology assessment method is that if the method is no longer supported or in use, citation of the results by the gateway developer becomes obsolete. Studies [22,23] and examination of previous reviews have shown that tools previously developed are no longer in use. Our study also found that the number of instruments has been reduced. It may be that people have begun to use instruments already in existence rather than to develop new ones. We examined citation of papers and Web addresses to estimate the current popularity of instruments on the basis that more popular technologies are more likely to survive. (In another field, the story of the VHS tape outliving the apparently technically better Betamax provides an example of the importance of "being popular.") Some of the instruments that we reviewed (eg, Kellogg), although they showed agreement with other instruments and were easy to use, may not survive because they have no critical mass of use.

The Ultimate Evaluation Instrument

We aimed to identify the best method for assessing websites for inclusion in a gateway on complementary medicine for breast cancer. No one tool seemed to be the answer. The three most-cited instruments on Google appeared to be DISCERN, HONcode, and NCCAM. HONcode does not seem to agree with the rankings produced by other instruments and seemed to have some quirks in its rankings. DISCERN seemed more difficult to apply than NCCAM, so if we chose one tool, it would be NCCAM. (This supports Walji's assertion that complementary medicine requires domain-specific criteria.) However, we think that the authors of instruments might benefit from merging their methods to produce one tool. This has recently been argued by Provost et al [77] in reporting the development of the WebMedQual scale. They argued that harmonization of Internet-based health information evaluative efforts would benefit all users and international researchers. They reviewed the literature on rating scales and identified 384 different items used by 26 scales. Four expert reviewers rated items, eliminated duplicates, and reworded or deleted items that were not clear, meaningful, or measurable, that were thought unimportant, too general, or vague, or that could not be feasibly ascertained by an experienced but nonmedical Internet user. They ended up with the following constructs: content (19 items), authority of source (18 items), design (19 items), accessibility and availability (6 items), links (4 items), user support (9 items), confidentiality and privacy (17 items), and e-commerce (6 items). They claimed that their scale, consisting of 8 categories, 8 subcategories, 95 items, and 3 supplemental items to assess website quality, was the first step toward a standard tool that would be easy to use. However, from our experience of using NCCAM and other instruments, we question whether an instrument requiring 98 items would be quick and easy to use.

A recently developed method of assessing websites containing health information, CLUE W (personal communication, Philippe Desjardins, Laval University, 2007), is designed to assess the clinical usefulness of information to a health professional. Interestingly, this instrument calculates the usefulness of a site from a formula that incorporates validity and relevance of the information on the site as well as the work required to use this information. This instrument has undergone an extensive development process involving many health professionals. With many instruments already in existence, it will be interesting to see how much attention this new assessment method will attract.

Another new method, FA4CT [87], published after our search, differs from the checklist approach by asking users to compare information they find with information on other sites; only if discordant information is found, a checklist (the CREDIBLE checklist) is used. This is referred to by the authors as a second generation educational model. Although this approach does not guarantee that information will be compared to a gold standard, it is claimed that this method of assessment is similar to the process that experts go through when searching for, and checking, the accuracy of information on the Internet. New methods such as FA4CT may make the checklist approach obsolete, but in the meantime, this study gives those developing gateways a practical guide as to which assessment instruments may be useful.

Acknowledgments

We thank Helen Cooke, Helen Seers, Karen Pilkington, and other colleagues at Penny Brohn Cancer Care for helpful comments. MB is funded as a Knowledge Transfer Partnership Associate by the Economic & Social Research Council (ESRC).

Conflicts of Interest

The Penny Brohn Centre is a charity providing complementary care to people affected by cancer.

Authors' Contributions

Matthew Breckons contributed to the study design, carried out all data collection, most data analysis, and co-wrote the paper. Ray Jones contributed to the study design, carried out some of the data analysis, and co-wrote the paper. Jenny Morris contributed to the study design and edited the paper. Janet Richardson contributed to the study design, edited the paper, secured the KTP partner, designed the KTP project, and is responsible for the academic management of the KTP project.

References

1. Satterlund MJ, McCaul KD, Sandgren AK. Information gathering over time by breast cancer patients. J Med Internet Res 2003 Aug 27;5(3):e15 [FREE Full text] [Medline: 14517106] [doi: 10.2196/jmir.5.3.e15]
2. Cole JI, Suman M, Schramm P, van Bel D, Lunn B, Maguire P, et al. The UCLA Internet Report: Surveying the Digital Future. Los Angeles, CA: UCLA Center for Communication Policy. URL: http://www.digitalcenter.org/pdf/InternetReportYearOne.pdf [accessed 2008 Jan 11] [WebCite Cache ID 5Qe4W9oGM]
3. Fox S, Rainie L. The online health care revolution: how the Web helps Americans take better care of themselves. Washington, DC: Pew Internet & American Life Project; 2003. URL: http://www.pewinternet.org/pdfs/PIP_Health_Report.pdf [accessed 2008 Jan 11] [WebCite Cache ID NVu]
4. Basch EM, Thaler HT, Shi W, Yakren S, Schrag D. Use of information resources by patients with cancer and their companions. Cancer 2004 Jun 1;100(11):2476-2483 [FREE Full text] [Medline: 15160355] [doi: 10.1002/cncr.20261]
5. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700 [FREE Full text] [Medline: 12020305] [doi: 10.1001/jama.287.20.2691]
6. Pérez-López FR, Pérez Roncero GR. Assessing the content and quality of information on the treatment of postmenopausal osteoporosis on the World Wide Web. Gynecol Endocrinol 2006 Dec;22(12):669-675. [Medline: 17162708] [doi: 10.1080/09513590601012603]
7. Butler L, Foster NE. Back pain online: a cross-sectional survey of the quality of web-based information on low back pain. Spine 2003 Feb 15;28(4):395-401. [Medline: 12590218] [doi: 10.1097/00007632-200302150-00017]
8. Hainer MI, Tsai N, Komura ST, Chiu CL. Fatal hepatorenal failure associated with hydrazine sulfate. Ann Intern Med 2000 Dec 5;133(11):877-880 [FREE Full text] [Medline: 11103057]
9. Pilkington K, Richardson J. Evidence-based complementary (and alternative) medicine on the Internet. He@lth Information on the Internet 2003;34(1):7-9 [FREE Full text] [WebCite Cache ID 5QFkVfd7J]
10. Schmidt K, Ernst E. Assessing websites on complementary and alternative medicine for cancer. Ann Oncol 2004 May;15(5):733-742 [FREE Full text] [Medline: 15111340] [doi: 10.1093/annonc/mdh174]
11. Weisbord SD, Soule JB, Kimmel PL. Poison on line--acute renal failure caused by oil of wormwood purchased through the Internet. N Engl J Med 1997 Sep 18;337(12):825-827. [Medline: 9297113] [doi: 10.1056/NEJM199709183371205]
12. Richardson J, Pilkington K. Complementary and alternative medicine evidence online for cancer. Br J Cancer Manage 2005;2(2):9-11.
13. Williams J, Clemens S, Oleinikova K, Tarvin K. The Skills for Life Survey: A National Needs and Impact Survey of Literacy, Numeracy and ICT Skills. London, UK: Department for Education and Skills; 2003. URL: http://www.dfes.gov.uk/research/data/uploadfiles/RB490.pdf [accessed 2008 Jan 11] [WebCite Cache ID 5UyFLblAo]
14. Nielsen-Bohlman L, Panzer AM, Kindig DA. Health Literacy: A Prescription to End Confusion. Washington, DC: The National Academies Press; 2004.
15. Ivanitskaya L, O'Boyle I, Casey AM. Health information literacy and competencies of information age students: results from the interactive online Research Readiness Self-Assessment (RRSA). J Med Internet Res 2006;8(2):e6 [FREE Full text] [Medline: 16867969] [doi: 10.2196/jmir.8.2.e6]
16. ; Commission of the European Communities, Brussels. eEurope 2002: Quality Criteria for Health Related Websites. J Med Internet Res 2002 Nov 29;4(3):E15 [FREE Full text] [Medline: 12554546] [doi: 10.2196/jmir.4.3.e15]
17. 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] [doi: 10.1136/bmj.324.7337.573]
18. Marshall LA, Williams D. Health information: does quality count for the consumer? How consumers evaluate the quality of health information materials across a variety of media. J Libr Inf Sci 2006;38(3):141-156. [doi: 10.1177/0961000606066575]

19. Schwartz KL, Roe T, Northrup J, Meza J, Seifeldin R, Neale AV. Family medicine patients' use of the Internet for health information: a MetroNet study. J Am Board Fam Med 2006;19(1):39-45 [FREE Full text] [Medline: 16492004]
20. Healthfinder. URL: http://www.healthfinder.gov/ [accessed 2008 Jan 15] [WebCite Cache ID 5QFjT3KK2]
21. Intute: Health and Life Sciences. URL: http://www.intute.ac.uk/healthandlifesciences/medicine [accessed 2008 Jan 15] [WebCite Cache ID 5QFjXWBeD]
22. Jadad AR, Gagliardi A. Rating health information on the Internet: navigating to knowledge or to Babel? JAMA 1998 Feb 25;279(8):611-614 [FREE Full text] [Medline: 9486757] [doi: 10.1001/jama.279.8.611]
23. Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. BMJ 2002 Mar 9;324(7337):569-573 [FREE Full text] [Medline: 11884320] [doi: 10.1136/bmj.324.7337.569]
24. Bernstam EV, Shelton DM, Walji M, Meric-Bernstam F. Instruments to assess the quality of health information on the World Wide Web: what can our patients actually use? Int J Med Inform 2005 Jan;74(1):13-19. [Medline: 15626632] [doi: 10.1016/j.ijmedinf.2004.10.001]
25. Kim P, Eng TR, Deering MJ, Maxfield A. Published criteria for evaluating health related web sites: review. BMJ 1999 Mar 6;318(7184):647-649 [FREE Full text] [Medline: 10066209]
26. ; Health Improvement Institute and Consumer Reports WebWatch. A Report on the Evaluation of Criteria Sets for Assessing Health Web Sites. Bethesda, MD: Health Improvement Institute; 2003. URL: http://www.consumerwebwatch.org/dynamic/health-report-assessing-health-sites-abstract.cfm [accessed 2008 Jan 15] [WebCite Cache ID 5QFjZytP2]
27. Web Feet Health Collection Criteria for Site Selection. Web Feet. URL: http://www.webfeetguides.com/criteria_WFforHealth.html [accessed 2008 Jan 15] [WebCite Cache ID 5QFjclQjk]
28. HON Code of Conduct (HONcode) for Medical and Health Web Sites. Health On the Net Foundation. URL: http://www.hon.ch/HONcode/Conduct.html [accessed 2008 Jan 15] [WebCite Cache ID 5QFjewHNN]
29. Health-Related Web Site Evaluation Form. 1998. Emory University Rollins School of Public Health. URL: http://www.sph.emory.edu/WELLNESS/instrument.html [accessed 2008 Jan 15] [WebCite Cache ID 5QFjhhpCL]
30. Web Site Evaluation Checklist. 1999. University of Michigan. URL: http://www-personal.umich.edu/~pfa/pro/courses/WebEvalNew.pdf [accessed 2008 Jan 15] [WebCite Cache ID 5QjvGTBtw]
31. Evaluation of Health Information on the Internet. 2001. Kellogg Library of Dalhousie University. URL: http://www.library.dal.ca/kellogg/internet/evaluate.htm [accessed 2008 Jan 15] [WebCite Cache ID 5QFjsWCM7]
32. The DISCERN Instrument. 1997. DISCERN. URL: http://www.discern.org.uk/discern_instrument.php [accessed 2008 Jan 15] [WebCite Cache ID 5UsQhfsFm]
33. 10 Things to Know About Evaluating Medical Resources on the Web. 2006. National Centre for Complementary and Alternative Medicine (NCCAM). URL: http://nccam.nih.gov/health/webresources/ [accessed 2008 Jan 15] [WebCite Cache ID 5QFjyAXZQ]
34. Anderson-Harper HM, Boley KO. Internet Enhances Patient Education. US Pharmacist. URL: http://www.uspharmacist.com/oldformat.asp?url=newlook/files/Feat/internet.cfm&pub_id=8&article_id=418 [accessed 2008 Jan 15] [WebCite Cache ID 5QFk18oIC]
35. LIDA: Minervation Validation Instrument for Health Care Web Sites. Minervation. URL: http://www.minervation.com/mod_product/LIDA/validation.aspx?o=1101 [accessed 2008 Jan 15] [WebCite Cache ID 5QFk3WJNe]
36. Nicoll LH. Quick and effective Website evaluation. Lippincotts Case Manag 2001;6(5):220-221. [Medline: 16398040] [doi: 10.1097/00129234-200109000-00006]
37. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351] [doi: 10.1001/jama.277.15.1244]
38. Sandvik H. Health information and interaction on the internet: a survey of female urinary incontinence. BMJ 1999 Jul 3;319(7201):29-32 [FREE Full text] [Medline: 10390457]
39. Quality Criteria for Health Related Websites. eEurope. URL: http://ec.europa.eu/information_society/eeurope/ehealth/quality/draft_guidelines/index_en.htm [accessed 2008 Jan 15] [WebCite Cache ID 5Tjz6RH2r]
40. Net Scoring Criteria to Assess the Quality of Health Internet Information. 2000. chu-rouen.fr. URL: http://www.chu-rouen.fr/netscoring/netscoringeng.html [accessed 2008 Jan 15] [WebCite Cache ID 5TjzZ74WM]
41. Criteria for Evaluating the Quality of Health Information on the Internet. Library of the Health Sciences-Urbana, University of Illinois at Chicago. URL: http://www.uic.edu/depts/lib/lhsu/resources/guides/web-evaluation.shtml [accessed 2008 Jan 15] [WebCite Cache ID 5TmqW5DbM]
42. Tomita M. Administration Design Quality Web Site Evaluation Method. 1999. steinhardt.nyu.edu. URL: http://steinhardt.nyu.edu/hepr/resources/online/adq.pdf [accessed 2008 Jan 15] [WebCite Cache ID 5TmqaGLWS]
43. Evaluating Websites. The University of Auckland. URL: http://www.library.auckland.ac.nz/instruct/evaluate.htm [accessed 2008 Jan 15] [WebCite Cache ID 5Tmqcr9OF]
44. Navigating the Health Care System: How to Evaluate Health Information on the Internet. 2005. Our Bodies Ourselves Health Resource Center. URL: http://www.ourbodiesourselves.org/book/excerpt.asp?id=39 [accessed 2008 Jan 15] [WebCite Cache ID 5Tmqg7AZf]

45. Rating Criteria and Excellence Awards. Growthhouse.org. URL: http://growthhouse.org/award.html [accessed 2008 Jan 15] [WebCite Cache ID 5TmqkeMQS]
46. Clean Bill of Health Award. University of Iowa. URL: http://www.lib.uiowa.edu/hardin/md/cbh.html [accessed 2008 Jan 15] [WebCite Cache ID 5TmqmCjYs]
47. ; Health Improvement Institute. Health Improvement Institute. URL: http://www.hii.org/documents/7004.pdf [accessed 2008 Jan 15] [WebCite Cache ID 5TmqoPwC4]
48. Best of the Web in Mental Health: Rating Guidelines. 2002. PsychCentral. URL: http://psychcentral.com/rateguid.htm [accessed 2008 Jan 15] [WebCite Cache ID 5ToFMU4Of]
49. Wyatt JC. Commentary: measuring quality and impact of the World Wide Web. BMJ 1997 Jun 28;314(7098):1879-1881 [FREE Full text] [Medline: 9224133]
50. Evaluating Internet Health Information: A Tutorial from the National Library of Medicine. Medline Plus. URL: http://www.nlm.nih.gov/medlineplus/webeval [accessed 2008 Jan 15] [WebCite Cache ID 5UsSny88m]
51. MedlinePlus Guide to Healthy Web Surfing. Medline Plus. URL: http://www.nlm.nih.gov/medlineplus/healthywebsurfing.html [accessed 2008 Jan 16] [WebCite Cache ID 5Utc5ATQv]
52. Taking Charge of Health Information. 2004. Health Insight. URL: http://www.health-insight.harvard.edu/guide.html [accessed 2008 Jan 16] [WebCite Cache ID 5ToGQSYZp]
53. How to Evaluate Health Information on the Internet: Questions and Answers. National Cancer Institute. URL: http://www.cancer.gov/cancertopics/factsheet/Information/internet [accessed 2008 Jan 16] [WebCite Cache ID 5UtcKNJcZ]
54. How to Find the Most Trustworthy Health Information on the Internet. Canadian Health Network. URL: http://www.canadian-health-network.ca/servlet/ContentServer?cid=1042668266229&pagename=CHN-RCS/Page/ShellStaticContentPageTemplate&c=Page&lang=En [accessed 2008 Jan 16] [WebCite Cache ID 5UtcWU74x]
55. Internet Detective. URL: http://www.vts.intute.ac.uk/detective/ [accessed 2008 Jan 16] [WebCite Cache ID 5TmrAbhzg]
56. ; Internet for Health and Well-Being. Intute. URL: http://www.vts.intute.ac.uk/acl/tutorial/wellbeing [accessed 2008 Jan 16] [WebCite Cache ID 5TmrE1f1Q]
57. Suggestions for Using the Internet to Find New Cancer Treatments. McGill Centre for Translational Research in Cancer. URL: http://www.mctrc.org/en/rp/suggestions.htm [accessed 2008 Jan 16] [WebCite Cache ID 5TmrLP4qF]
58. How to Judge the Quality of a Web Site. 2006. Judge: Web Sites for Health. URL: http://www.judgehealth.org.uk/how_to_judge.htm [accessed 2008 Jan 16] [WebCite Cache ID 5TmrQJOBZ]
59. ; Intute: Health & Life Sciences. Intute: Health and Life Sciences Evaluation Guidelines. Intute. URL: http://www.intute.ac.uk/healthandlifesciences/BIOME_Evaluation_Guidelines.doc [accessed 2008 Jan 16] [WebCite Cache ID 5TmrTGBVl]
60. Best Practice Web Assessments. Evaluation Criteria. 2005. e-Gov Watch. URL: http://www.e-govwatch.org.nz/criteria/ [accessed 2008 Jan 16] [WebCite Cache ID 5TmrVgSQe]
61. Fahey DK, Weinberg J. LASIK complications and the Internet: is the public being misled? J Med Internet Res 2003 Mar 31;5(1):e2. [Medline: 12746207] [doi: 10.2196/jmir.5.1.e2]
62. Quality Standards for Medical Publishing on the Web. Bhia.org. URL: http://www.bhia.org/ [accessed 2008 Jan 16] [WebCite Cache ID 5UzNg8NI2]
63. Beckner WM, Berman BM. Complementary Therapies on the Internet. London, UK: Churchill Livingstone; 2003:45-53.
64. Complementary Therapies: A Brief Guide. Breast Cancer Care UK. URL: http://www.breastcancercare.org.uk/content.php?page_id=1013 [accessed 2008 Jan 17] [WebCite Cache ID 5UuhdujER]
65. ; Breast Cancer Haven. Links. Breast Cancer Haven UK. URL: http://www.breastcancerhaven.org.uk/policy.php?parent=87 [accessed 2008 Jan 17] [WebCite Cache ID 5Uui75ssH]
66. Research Into Living With Breast Cancer. Cancer Research UK. URL: http://www.cancerhelp.org.uk/help/default.asp?page=10251 [accessed 2008 Jan 17] [WebCite Cache ID 5UuiQBERt]
67. Breast Cancer Treatment. Imaginis. URL: http://imaginis.com/breasthealth/alternative.asp [accessed 2008 Jan 17] [WebCite Cache ID 5UujSuUI0]
68. Complementary/Integrative Medicine. MD Anderson Cancer Center. URL: http://www.mdanderson.org/departments/cimer/ [accessed 2008 Jan 17] [WebCite Cache ID 5UujjfS4p]
69. Complementary Therapies. Cancer Treatment Centers of America. URL: http://www.cancercenter.com/complementary-alternative-medicine.cfm [accessed 2008 Jan 17] [WebCite Cache ID 5UujsR5nN]
70. Complementary Therapies for Dealing With Secondary Breast Cancer. Cancerbackup. URL: http://www.cancerbackup.org.uk/Cancertype/Breastsecondary/Livingwithcancer/Complementarytherapies [accessed 2008 Jan 17] [WebCite Cache ID 5UukABrZg]
71. Heart Spring. URL: http://heartspring.net/ [accessed 2008 Jan 17] [WebCite Cache ID 5UukN4yMq]
72. Issels Treatment. URL: http://www.issels.com/?c1=ppc&source=google&kw=breastcancertreatment [accessed 2008 Jan 17] [WebCite Cache ID 5Uuko8pui]
73. Alternative Cancer. URL: http://www.alternative-cancer.net/index.html [accessed 2008 Jan 17] [WebCite Cache ID 5UukyJZRy]
74. Cancer Treatment: Complementary and Alternative Medicine. MedicineNet. URL: http://www.medicinenet.com/script/main/art.asp?articlekey=63338 [accessed 2008 Jan 17] [WebCite Cache ID 5UulRY1lM]

75. Focus on Health. Elbee Global. URL: http://www.elbeeglobal.com/ [accessed 2008 Jan 17] [WebCite Cache ID 5UulpdFlm]
76. Bernstam EV, Sagaram S, Walji M, Johnson CW, Meric-Bernstam F. Usability of quality measures for online health information: Can commonly used technical quality criteria be reliably assessed? Int J Med Inform 2005 Aug;74(7-8):675-683. [Medline: 16043090] [doi: 10.1016/j.ijmedinf.2005.02.002]
77. Provost M, Koompalum D, Dong D, Martin BC. The initial development of the WebMedQual scale: domain assessment of the construct of quality of health web sites. Int J Med Inform 2006 Jan;75(1):42-57. [Medline: 16169770] [doi: 10.1016/j.ijmedinf.2005.07.034]
78. Impicciatore P, Pandolfini C, Casella N, Bonati M. Reliability of health information for the public on the World Wide Web: systematic survey of advice on managing fever in children at home. BMJ 1997 Jun 28;314(7098):1875-1879 [FREE Full text] [Medline: 9224132]
79. Greene DL, Appel AJ, Reinert SE, Palumbo MA. Lumbar disc herniation: evaluation of information on the internet. Spine 2005 Apr 1;30(7):826-829. [Medline: 15803087] [doi: 10.1097/01.brs.0000157754.98023.cd]
80. Selman TJ, Prakash T, Khan KS. Quality of health information for cervical cancer treatment on the internet. BMC Womens Health 2006;6:9 [FREE Full text] [Medline: 16787534] [doi: 10.1186/1472-6874-6-9]
81. Pandolfini C, Impicciatore P, Bonati M. Parents on the web: risks for quality management of cough in children. Pediatrics 2000 Jan;105(1):e1 [FREE Full text] [Medline: 10617738] [doi: 10.1542/peds.105.1.e1]
82. Walji M, Sagaram S, Sagaram D, Meric-Bernstam F, Johnson C, Mirza NQ, et al. Efficacy of quality criteria to identify potentially harmful information: a cross-sectional survey of complementary and alternative medicine web sites. J Med Internet Res 2004 Jun 29;6(2):e21 [FREE Full text] [Medline: 15249270] [doi: 10.2196/jmir.6.2.e21]
83. Cancer Specialist Library. National Library for Health. URL: http://www.library.nhs.uk/cancer/ [accessed 2008 Jan 16] [WebCite Cache ID 5TmreqgKD]
84. National Center for Complementary and Alternative Medicine (NCCAM). URL: http://nccam.nih.gov/ [accessed 2008 Jan 16] [WebCite Cache ID 5TmrchXTB]
85. School of Integrated Health. Complementary and Alternative Medicine Evidence OnLine (CAMEOL). The Research Council for Complementary Medicine. URL: http://www.rccm.org.uk/cameol/ [accessed 2008 Jan 16] [WebCite Cache ID 5TmranyZL]
86. Ademiluyi G, Rees CE, Sheard CE. Evaluating the reliability and validity of three tools to assess the quality of health information on the Internet. Patient Educ Couns 2003 Jun;50(2):151-155. [Medline: 12781930]
87. Eysenbach G, Thomson M. The FA4CT algorithm: a new model and tool for consumers to assess and filter health information on the Internet. Medinfo 2007;12(Pt 1):142-146 [FREE Full text] [Medline: 17911695]

Abbreviations

HIICRW: Health Improvement Institute and Consumer Reports WebWatch
HONcode: HON Code of Conduct

submitted 01.08.07; peer-reviewed by E Bernstam; comments to author 13.11.07; revised version received 03.12.07; accepted 04.01.08; published 22.01.08.

Please cite as:
Breckons M, Jones R, Morris J, Richardson J
What Do Evaluation Instruments Tell Us About the Quality of Complementary Medicine Information on the Internet?
J Med Internet Res 2008;10(1):e3
URL: http://www.jmir.org/2008/1/e3/
doi:10.2196/jmir.961
PMID:18244894

Original Paper

Evaluation of Influenza Prevention in the Workplace Using a Personally Controlled Health Record: Randomized Controlled Trial

Florence T Bourgeois[1,2], MD, MPH; William W Simons[3], MS; Karen Olson[1,3], PhD; John S Brownstein[1,2,3], PhD; Kenneth D Mandl[1,2,3], MD, MPH

[1]Division of Emergency Medicine, Children's Hospital Boston, Boston, MA, USA

[2]Department of Pediatrics, Harvard Medical School, Boston, MA, USA

[3]Children's Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology, Boston, MA, USA

Corresponding Author:
Florence T Bourgeois, MD, MPH
Division of Emergency Medicine
Children's Hospital Boston
300 Longwood Avenue
Boston, MA 02115
USA
Phone: +1 857 218 4555
Fax: +1 617 730 0699
Email: florence.bourgeois@childrens.harvard.edu

Abstract

Background: Personally controlled health records (PCHRs) are accessible over the Internet and allow individuals to maintain and manage a secure copy of their medical data. These records provide a new opportunity to provide customized health recommendations to individuals based on their record content. Health promotion programs using PCHRs can potentially be used in a variety of settings and target a large range of health issues.

Objectives: The aim was to assess the value of a PCHR in an employee health promotion program for improving knowledge, beliefs, and behavior around influenza prevention.

Methods: We evaluated a PCHR-based employee health promotion program using a randomized controlled trial design. Employees at Hewlett Packard work sites who reported reliable Internet access and email use at least once every 2 days were recruited for participation. PCHRs were provided to all participants for survey administration, and tailored, targeted health messages on influenza illness and prevention were delivered to participants in the intervention group. Participants in the control group received messages addressing cardiovascular health and sun protection. The main outcome measure was improvement in knowledge, beliefs, and behavior around influenza prevention. Secondary outcomes were influenza vaccine rates among household members, the impact of cardiovascular health and sun protection messages on the control group, and the usability and utility of the PCHR-based program for employees.

Results: The intervention did not have a statistically significant effect on the influenza knowledge elements we assessed but did impact certain beliefs surrounding influenza. Participants in the intervention group were more likely to believe that the influenza vaccine was effective (OR = 5.6; 95% CI = 1.7-18.5), that there were actions they could take to prevent the flu (OR = 3.2; 95% CI = 1.1-9.2), and that the influenza vaccine was unlikely to cause a severe reaction (OR = 4.4; 95% CI = 1.3-15.3). Immunization rates did not differ between the intervention and control groups. However, participants in the intervention group were more likely to stay home during an infectious respiratory illness compared with participants in the control group (39% [16/41] vs 14% [5/35], respectively; $P = .02$). The program also succeeded in improving recognition of the signs of heart attack and stroke among participants in the control group. Overall, 78% of participants rated the PCHR as "extremely/very" easy to use, and 73% responded that they would be "extremely/very" likely to participate again in a PCHR-based health promotion system such as this one.

Conclusions: With a small sample size, this study identified a modest impact of a PCHR-based employee health program on influenza prevention and control. Employees found the PCHR acceptable and easy to use, suggesting that it should be explored as a common medium for health promotion in the workplace.

Trial Registration: ClinicalTrials.gov NCT00142077

(J Med Internet Res 2008;10(1):e5) doi:10.2196/jmir.984

KEYWORDS

Randomized controlled trial; personally controlled health record; Web-based; employee health program; influenza

Introduction

Yearly influenza outbreaks are a prime example of a public health problem with well-developed surveillance methods and evidence-based programs for prevention but poor compliance with health protection guidelines [1]. Fifty-five million adults aged 18 to 64 years are infected with influenza every year, with 200 million days of restricted activity, 70 million days of work absenteeism, and 18 million visits to health care providers [2]. Influenza vaccination among healthy, working adults has been shown to be highly effective, resulting in a 25% reduction in any episode of upper respiratory illness, a 43% decrease in days of work missed due to respiratory illness, and 44% fewer visits to physicians' offices for respiratory illnesses when compared to unvaccinated adults [3]. Nonetheless, vaccination rates are only 18% among healthy adults 18 to 49 years of age and 46% among those with high-risk conditions 50-64 years of age [4].

Personally controlled health records (PCHRs) [5] are a subset of personal health records [1,6] and enable an individual to assemble, maintain, and manage a secure copy of his or her medical data [7]. PCHRs are designed based on the principle that patients have the right to own and manage copies of their own medical histories, and they provide a virtual medical home with modalities for communication among patients, clinicians, and health authorities. PCHRs present a new opportunity to bridge the gap between public health research and action to improve the health of individuals. We explored the use of a PCHR as a vehicle for the delivery of customized health promotion messages in which individuals received information and recommendations based on their record content. This approach to health communication enables rapid, tailored, and targeted delivery of health care recommendations to individuals. Tailored communication has previously been shown to be superior to generic, population-based recommendations in achieving patient compliance [8] and can easily be implemented with PCHRs.

We report an evaluation of a PCHR-based employee health promotion program using a randomized controlled trial design. The principle objective was to assess the use of the PCHR to improve knowledge, beliefs, and behavior surrounding influenza prevention. There were three secondary objectives. The first was to assess the effect of electronic messages delivered through the PCHR on influenza vaccine rates among household members, the second was to assess the impact of messages addressing cardiovascular health and sun protection on the knowledge and behavior among participants in the control group, and the third was to evaluate the usability and utility of the PCHR-based program for employees.

Methods

Design and Participants

Using a randomized controlled trial design (ClinicalTrials.gov: NCT00142077), we evaluated an electronic PCHR system to modify knowledge, beliefs, and behavior around influenza. Participants were recruited from eight Hewlett Packard Corporation work sites in the northeastern United States in the fall of 2005. Employees at the research sites were recruited with two emails sent to their work email address by the company's human resources department. The emails contained study information and invited potential participants to complete a brief set of questions to assess eligibility. Eligible volunteers were 18 years of age or older, comfortable reading and writing in English, part-time or full-time employees of the company and had reliable Internet access at work, school, or home and used email at least once every 2 days. In addition, participants could not have a history of a severe reaction to influenza vaccine or severe allergy to chicken eggs, since both of these conditions contraindicate use of the influenza vaccine. Enrollment was initially planned for October 2005; however, just prior to the original recruitment period, several Hewlett Packard work sites were closed and employees were relocated or laid off at several other sites. Therefore, the study began in November 2005 and our recruitment pool was smaller than anticipated. All participants electronically provided informed consent prior to study initiation. The study was approved by the Committee on Clinical Investigation at Children's Hospital Boston, Boston, MA, USA.

Assignment to the intervention and control groups was performed at the level of the corporation work site in order to prevent employees at the same site from sharing information about the trial, including recommendations provided in the health messages. Prior to study initiation, we created two groups with four sites in each such that the number of employees in each arm was evenly distributed. The two groups were then randomly assigned to the intervention and control arms by a person unfamiliar with the details of the work sites. Participants in the study were informed that the study was to evaluate health promotion using a PCHR with an electronic messaging system and were masked as to whether they were in the intervention or control groups.

Interventions

We used a PCHR system called PING [9-11] (new versions are called Indivo [12]), which is built to open standards on a flexible XML data model and is accessible over the Web. PING is designed to enable patients to own complete, secure copies of their medical record and to integrate information over time and across sites of care [5,11]. In this investigation, we tested the survey, decision support, and health messaging features of

PING, and records did not include any health information beyond the data provided by subjects for this study. Enrolled subjects completed online health risk assessment surveys, the responses to which drove the decision support system to generate and send tailored health messages for participants in the intervention group. These messages were sent to participants' PING record inbox, and participants were simultaneously notified with a standard, plain-text email instructing them to visit and log on to their PING record to review the message (Multimedia Appendix 1: PING Record Welcome Screen and Inbox).

Data Collection for PCHR

Participants in the intervention and control groups completed three types of survey. The first was a baseline survey that was posted in their PING record immediately after registration was completed (Multimedia Appendix 2: Enrollment Survey). This survey collected demographic data; information on medical history; health-related behaviors; influenza risk factors; knowledge, beliefs, and behavior around influenza; and information related to Internet use. Information was also collected on household members, including their age; gender; attendance at work, school, or daycare; and behaviors and risk factors related to influenza. The baseline survey administered to the control group contained additional questions addressing routine health and knowledge and behaviors regarding cardiovascular health and sun protection.

The second survey was a biweekly survey consisting of a brief set of questions that was administered approximately every 2 weeks (Multimedia Appendix 3: Biweekly Survey). A total of seven of these surveys were administered between December 1, 2005, and March 1, 2006. Information was collected on recent respiratory illnesses in participants and household members, including duration of symptoms, missed work or school days, medication use, and health care utilization, and an update was obtained on their influenza vaccine status. Biweekly surveys for the control group included additional questions on routine health care use and recent gastrointestinal or other illness.

The third survey was an exit survey administered at the end of the study, 2 weeks after the last biweekly survey. It contained the same questions on influenza knowledge, beliefs, and behavior as the baseline survey, as well as questions to evaluate the electronic interface of the application, its usability, the content of the questions, and the overall utility of the PCHR-based program to participants. The survey administered to the control group additionally contained the same questions on knowledge and behaviors regarding cardiovascular health and sun protection administered in the baseline survey.

Health Messages

Participants in the intervention group received different types of influenza-related health messages throughout the study period. Some of these were personalized based on the information provided in the baseline and biweekly surveys and were posted in the record after a participant completed one of these surveys. Messages were tailored to include advice for all household members, to identify individuals at high risk for influenza-related complications, and to provide information on respiratory illnesses if a participant or household member became ill with a respiratory infection. The health messages were also tailored based on the home addresses of participants to advise them of influenza activity in their area. Other messages contained general information and were provided on a weekly or monthly basis. The content of the health messages was regularly monitored throughout the study period to ensure that proper messages were being generated and transmitted to participants.

There were five types of health message:

1. Vaccine reminders: If participants indicated that they or a household member eligible for the influenza vaccine were not yet vaccinated, a message was generated urging them to receive the vaccine. The message contained basic information on the influenza vaccine and identified any household members who were at high risk for influenza-related complications or severe disease based on recommendations by the Advisory Committee on Immunization Practices (ACIP) [13]. Figure 1 provides an example of such a personalized health message.
2. Respiratory illness advice: Information that a participant or household member had recently contracted a respiratory illness prompted a health message with advice on the treatment and prevention of respiratory illnesses (Multimedia Appendix 4: Sample Health Message). Participants were encouraged to stay home from work when ill, and guidelines were provided on when to contact a physician.
3. Influenza alerts: Based on surveillance information provided by the Centers for Disease Control and Prevention (CDC) on mortality related to influenza and pneumonia [14], weekly messages were sent to participants residing in areas with increased rates of death attributable to pneumonia and influenza. These messages alerted participants to the increase in influenza activity in their area and contained information on preventing influenza transmission.
4. Weekly influenza risk maps: Every week, participants received a map displaying areas of low, moderate, and high influenza activity in the northeastern United States. These maps were based on the weekly CDC surveillance of pneumonia and influenza [14] and kept participants informed of the spread of influenza in their region. Figure 2 is an example of such a map.
5. Monthly bulletins: Once a month, a message was sent with educational information about different aspects of influenza. A total of four such messages were sent, describing methods of influenza transmission and prevention, symptoms of influenza illness, influenza vaccine and its risks, and treatment options for influenza illness.

Figure 1. Sample health message: vaccine reminder (In this example, Alice and Bob are household members of the study participant.)

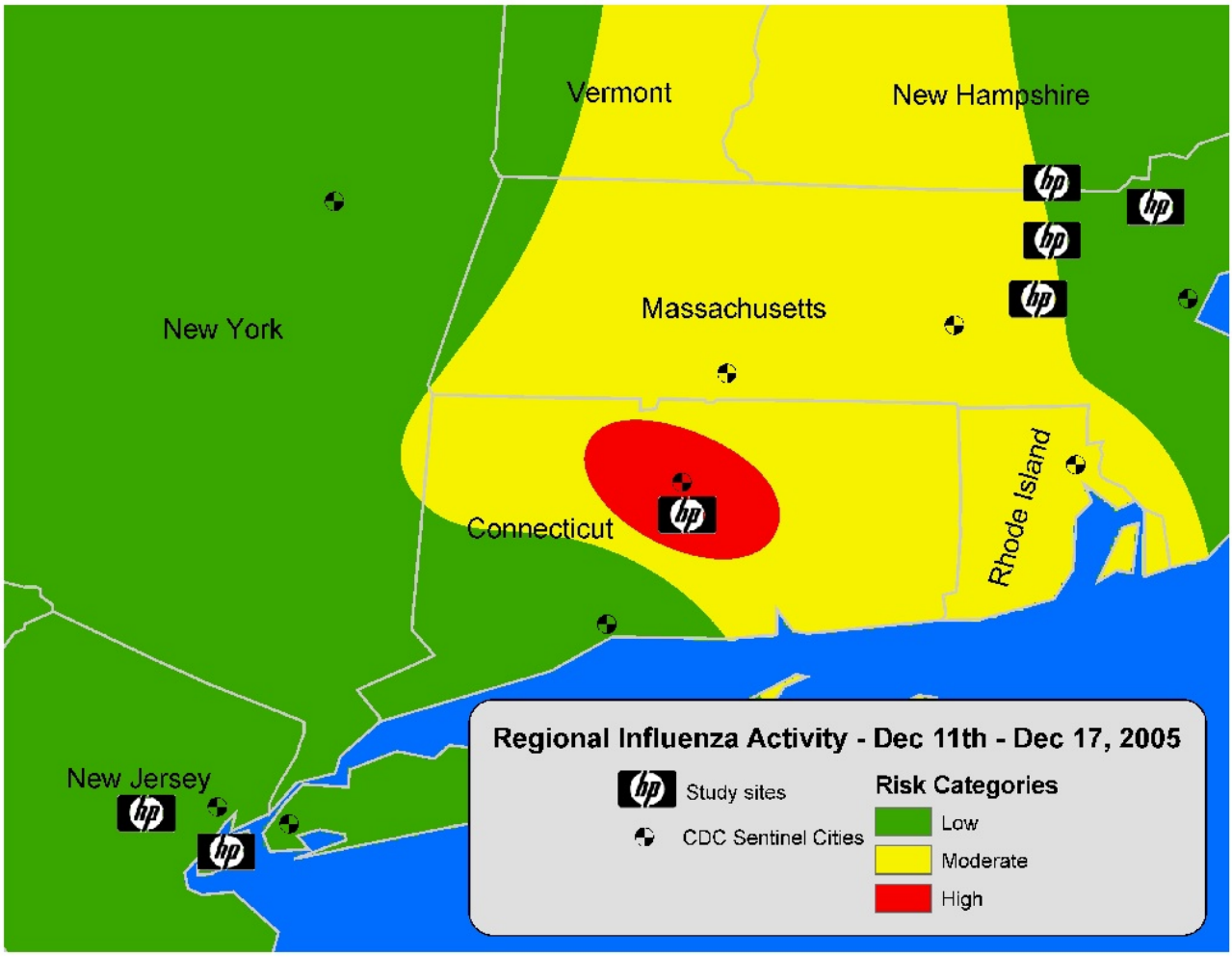
Get the flu vaccine! This is the most effective way to prevent the flu. According to the information you have provided, not everyone in your household has received the flu vaccine this year. **Alice and Bob still need to get the flu vaccine.** It is especially important that **Bob** get the vaccine since he is **at high risk for serious complications from the flu.**

Everyone should get the flu vaccine every year in order to prevent getting the flu. October and November are the best times to get vaccinated, but you can still get vaccinated in December and later. The flu season can last as late as May.

Remember that ***everyone*** can get the flu and the vaccine is ***very effective*** at preventing the flu. ***The vaccine cannot give you the flu.***

For more information on the flu vaccine, including who should get the vaccine and what the common side effects are, go to www.cdc.gov/flu/protect/keyfacts.htm.

Figure 2. Sample health message: weekly influenza risk map (Risk categories were derived from data provided by the CDC on weekly mortality from pneumonia and influenza [9].)



Participants in the control group received a monthly bulletin on cardiovascular health and sun protection (Multimedia Appendix 5: Control Group Monthly Bulletin). Information was selected based on Healthy People 2010 [15] objectives, which aim to reduce high-risk behaviors and improve the use of preventive services. Participants in the control group received neither personalized health messages nor information on influenza. Four bulletins were sent and provided information on cardiovascular disease, stroke, skin cancer and sun protection, and guidelines for a healthy diet.

Outcomes

The primary outcome was change in knowledge, beliefs, and behavior surrounding influenza prevention. Change in knowledge was assessed using a set of nine questions in the baseline and exit surveys addressing influenza transmission and prevention, influenza illness, and the influenza vaccine. Change in beliefs was measured with a set of six questions on influenza illness and vaccine, administered in the baseline and exit surveys. Measurements of behavior change consisted of the rate of influenza vaccination, the rate of work attendance despite a respiratory illness, and responses to two questions in the baseline and exit surveys on hand hygiene and cough etiquette.

Secondary outcomes included the rate of influenza vaccination among household members and changes in knowledge and behavior regarding cardiovascular health and sun protection, measured using nine questions in the baseline and exit surveys administered to the control group. Finally, the usability and utility of the PCHR-based program were assessed by means of 12 questions in the exit survey, as well as survey completion rates and mean days to survey completion.

Statistical Methods

Logistic regression models were used to analyze the changes in responses for the questions on knowledge and behavior surrounding influenza. The models controlled for baseline responses in the initial survey. A variable for participant work sites was tested in the models to control for clustering. This variable was not significant in any of the analyses and was excluded from the final models. Immunization rates were compared using chi-square analysis. Rates for missed work during an illness were examined using the SAS v9.1 (SAS Institute, Inc, Cary, NC, USA) PROC GENMOD procedure in order to control for correlated responses from participants with more than one illness. The knowledge and behavior questions on cardiovascular risk and sun protection for the control group were analyzed with the McNemar test. Assessment of the usability and utility of the program was performed through examination of the responses to the questions on user experience and calculation of completion rates and mean days to survey completion.

Results

Participation and Retention

Participant flow is shown in Figure 3. We recruited participants during a 4-week period between November 10 and December 7, 2005. Of the 3540 employees at the eight work sites, 144 employees registered for the study and 125 completed the baseline survey. Of these, 119 (95%) completed between one and seven biweekly surveys, and 99 (79%) completed the exit survey. The baseline characteristics of the intervention and control groups are shown in Table 1. Only the gender distribution differed between the two groups.

Figure 3. Flow diagram of study participation

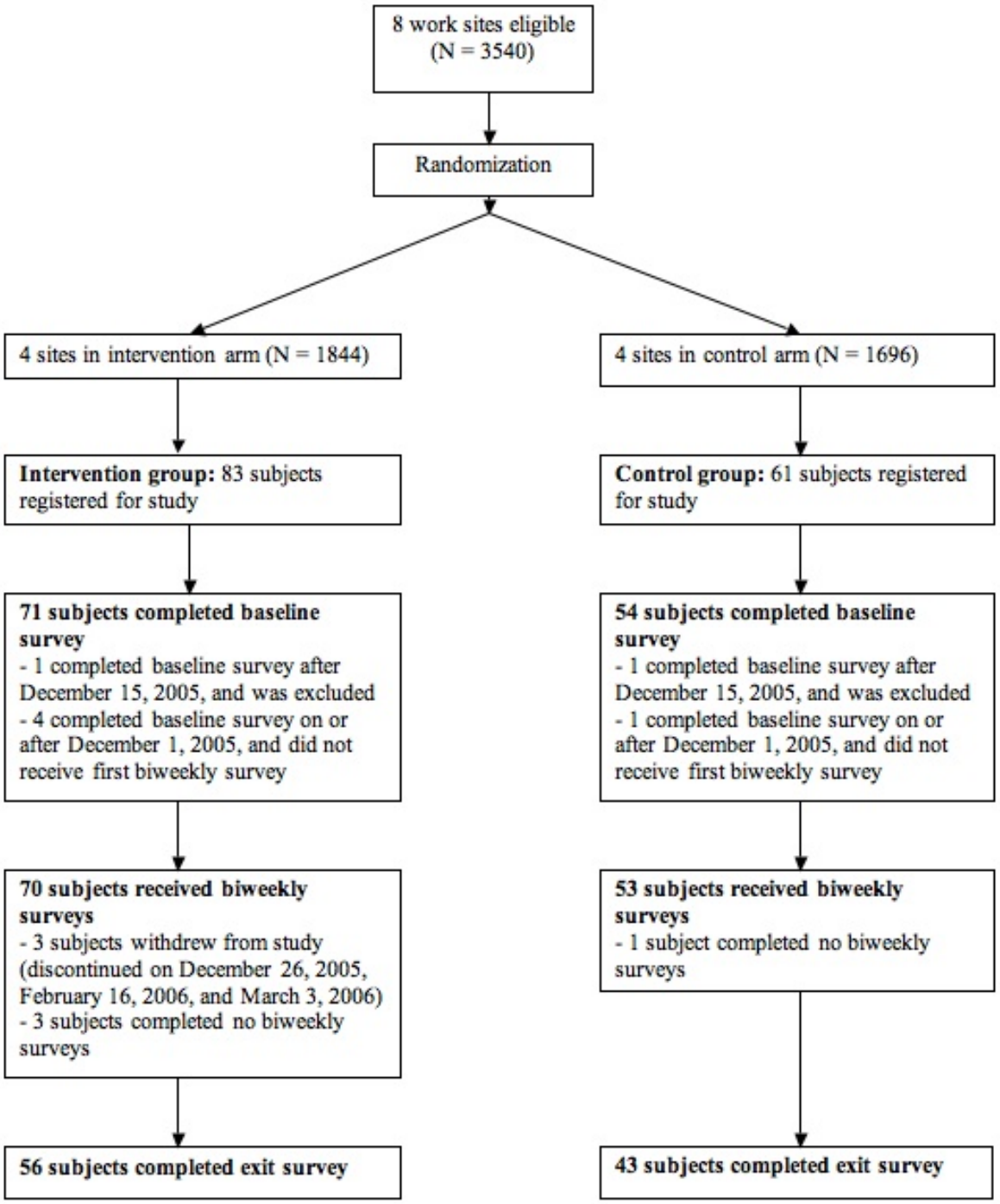


Table 1. Baseline characteristics of participants

Characteristic	Intervention (N = 71)	Control (N = 54)	*P* Value
Number of female participants (%)	41 (58)	20 (37)	.02
Mean age in years (SD)	46.4 (8.6)	46.9 (9.4)*	.47
Number at increased risk of complications† (%)	10 (14)	9 (17)	.69
Number who received influenza vaccine during previous flu season (%)	19 (27)	13 (24)	.73
Number who received influenza vaccine during current flu season prior to study start (%)	14 (20)	9 (17)	.66

*One control subject excluded due to incorrect input of birth date.

†Based on recommendations by the Advisory Committee on Immunization Practices (ACIP) [13].

Of the 125 participants completing the baseline survey, two were excluded because they completed it too late (on March 16, 2006, and April 16, 2006). A total of 99 participants completed the exit survey and were included in the analyses examining changes in knowledge, beliefs, and behaviors. There were 123 participants who received biweekly surveys, among which four did not complete any of the biweekly surveys (or the exit survey) and were excluded from the analyses of vaccination rates and work attendance rates while ill. Among the control group, there were 43 participants who completed both the baseline and exit survey and were included in the analysis examining changes in knowledge and behavior regarding cardiovascular risk and sun protection. For the assessment of the usability and utility of the PCHR, we analyzed the responses to the questions on user experience in the exit survey completed by 99 participants, as well as the completion rates and times to completion of 123 participants for the enrollment survey, 119 participants for the biweekly surveys, and 99 participants for the exit survey.

Outcomes and Estimation

Improvement in Knowledge, Beliefs, and Behavior Surrounding Influenza

Table 2 summarizes responses to survey questions evaluating knowledge, beliefs, and behaviors surrounding influenza. The intervention did not have a statistically significant effect on the knowledge elements we assessed. However, it did have a significant effect on certain beliefs surrounding influenza. At the end of the study, participants in the intervention group were more likely to believe that the influenza vaccine was effective (OR = 5.6; 95% CI = 1.7-18.5), that there were actions they could take to prevent the flu (OR = 3.2; 95% CI = 1.1-9.2), and that the influenza vaccine was unlikely to cause a severe reaction (OR = 4.4; 95% CI = 1.3-15.3). The intervention was not demonstrably effective in changing people's beliefs that they should be immunized, that influenza illness is a moderately to extremely serious illness, or that immunization can help prevent influenza in other people. The two questions addressing hand hygiene and cough etiquette did not show any changes in behavior among the intervention group.

Table 2. Effect of intervention on knowledge, beliefs, and behavior regarding influenza

	Intervention (N = 56)		Control (N = 43)		OR (95% CI)[*,†]	*P* Value[†]
	Baseline Survey	Completion Survey	Baseline Survey	Completion Survey		
Knowledge	**Participants Responding Correctly, N (%)**					
Q1. Infection: contacts	55 (98)	54 (96)	42 (98)	41 (95)	1.3 (0.2-9.8)	.78
Q2. Infection: unhealthy behaviors	20 (36)	20 (36)	21 (49)	19 (44)	0.9 (0.3-2.3)	.81
Q3. Infection: cold conditions	49 (88)	47 (84)	38 (88)	37 (86)	0.9 (0.3-2.8)	.79
Q4. Infection: untreated illness	45 (80)	35 (62)	32 (74)	26 (60)	1.0 (0.4-2.4)	.91
Q5. Influenza vaccine	33 (59)	40 (71)	19 (44)	24 (56)	0.6 (0.2-1.8)	.38
Q6. Hand hygiene	56 (100)	55 (98)	43 (100)	40 (93)	4.1 (0.4-41.1)	.23
Q7. Cough etiquette	44 (79)	46 (82)	28 (65)	35 (81)	0.7 (0.2-2.3)	.56
Q8. Hand cleaners	9 (16)	13 (23)	6 (14)	7 (16)	1.6 (0.5-5.2)	.42
Q9. Work attendance	45 (80)	47 (84)	34 (79)	31 (72)	2.3 (0.8-6.7)	.14
Beliefs	**Participants Responding in the Affirmative,[‡] N (%)**					
Q1. Vaccine effectiveness	43 (77)	49 (88)	29 (67)	26 (60)	**5.6 (1.7-18.5)**	**.003**
Q2. Vaccine eligibility	43 (77)	44 (79)	28 (65)	28 (65)	1.7 (0.5-6.2)	.41
Q3. Influenza prevention	36 (64)	49 (88)	30 (70)	30 (70)	**3.2 (1.1-9.2)**	**.03**
Q4. Influenza illness	44 (79)	47 (84)	38 (88)	37(86)	1.2 (0.3-4.1)	.80
Q5. Vaccine benefits	40 (71)	44 (79)	30 (70)	33 (77)	1.1 (0.3-3.7)	.89
Q6. Vaccine reactions	36 (64)	45 (80)	30 (70)	28 (65)	**4.4 (1.3-15.3)**	**.02**
Behavior	**Participants Responding in the Affirmative,[‡] N (%)**					
Q1.a. Hand hygiene	48 (86)	50 (89)	40 (93)	40 (93)	0.9 (0.2-4.4)	.88
Q1.b. Hand hygiene	37 (66)	47 (84)	28 (65)	35 (81)	1.2 (0.4-3.8)	.75
Q1.c. Hand hygiene	41 (73)	48 (86)	38 (88)	37 (86)	1.9 (0.5-7.6)	.36
Q2.a. Cough etiquette	46 (82)	38 (68)	24 (56)	31 (72)	0.7 (0.3-1.6)	.37
Q2.b. Cough etiquette	37 (66)	52 (93)	28 (65)	37 (86)	2.3 (0.5-9.6)	.27
Q2.c. Cough etiquette	30 (54)	28 (50)	27 (63)	22 (51)	1.0 (0.4-2.5)	.93
Q2.d. Cough etiquette	49 (88)	55 (98)	38 (88)	39 (91)	5.7 (0.6-53.4)	.13
Q2.e. Cough etiquette	31 (55)	44 (79)	23 (53)	30 (70)	1.8 (0.6-5.1)	.30
Q2.f. Cough etiquette	19 (34)	33 (59)	16 (37)	25 (58)	1.1 (0.5-2.7)	.81

[*]Logistic regression model controlling for baseline responses.

[†]Statistically significant effects indicated in bold.

[‡]Refers to responses indicating beliefs or behaviors conducive to preventing influenza illness.

We also examined the rate of influenza immunization among participants during the study period and the rate of work attendance despite a respiratory illness. We did not detect a significant difference in the rate of immunization between the intervention and control groups (24% [13/54] vs 19% [8/43], respectively; $P = .50$). There were a total of 76 participants who reported at least one respiratory illness during the study period, with 21 missing work as a result of an illness. A higher proportion of participants in the intervention group (39%, 16/41) stayed home during an illness compared with participants in the control group (14%, 5/35; $P = .02$).

Vaccination Rate Among Household Members

Participants provided information on 160 household members, among which 158 were eligible for the influenza vaccine (two were younger than 6 months at the start of the study and therefore not eligible): 15.8% (13/82) of household members in the intervention group and 9.2% (7/76) in the control group received the influenza vaccine during the study period ($P = .21$).

Changes in Knowledge and Behavior in the Control Group

Table 3 shows responses to the survey questions evaluating knowledge and behavior around cardiovascular health and sun protection in the control group. At the end of the study,

participants in the control group were more likely to recognize "pain or discomfort in the jaw, neck, or back" and "feeling weak, lightheaded, or faint" as signs of a heart attack and "sudden trouble seeing in one or both eyes" and "severe headache with no known cause" as signs of a stroke. The intervention did not significantly affect the other knowledge elements tested or the proportion of subjects taking medication for their high blood pressure, having their cholesterol checked, or taking measures toward sun protection.

Table 3. Changes in knowledge and behavior regarding cardiovascular health and sun protection in the control group (N = 43)

	Baseline Survey	Completion Survey	McNemar Test, *P* Value[*]
Knowledge	**Participants Responding Correctly, N (%)**		
Q1.a. Heart attack recognition	18 (42)	34 (79)	**.007**
Q1.b. Heart attack recognition	23 (53)	33 (77)	**.02**
Q1.c. Heart attack recognition	41 (95)	43 (100)	N/A
Q1.d. Heart attack recognition	19 (44)	14 (33)	.16
Q1.e. Heart attack recognition	35 (81)	38 (88)	.18
Q1.f. Heart attack recognition	35 (81)	39 (91)	.20
Q2.a. Stroke recognition	39 (91)	42 (98)	.18
Q2.b. Stroke recognition	40 (93)	43 (100)	N/A
Q2.c. Stroke recognition	33 (77)	39 (91)	**.01**
Q2.d. Stroke recognition	21 (49)	19 (44)	.48
Q2.e. Stroke recognition	37 (86)	41 (95)	.10
Q2.f. Stroke recognition	25 (58)	33 (77)	**.02**
Q3. Interventions	40 (93)	43 (100)	N/A
Behavior	**Participants Responding in the Affirmative,[†] N (%)**		
Q1. High blood pressure	3 (33)	3 (33)	1.0
Q2. Cholesterol monitoring	40 (93)	40 (93)	1.0
Q3.a. Heart disease prevention	33 (77)	32 (74)	.74
Q3.b. Heart disease prevention	34 (79)	33 (77)	.74
Q3.c. Heart disease prevention	34 (79)	31 (72)	.32
Q4.a. Sun protection	20 (47)	22 (51)	.59
Q4.b. Sun protection	28 (65)	32 (74)	.21
Q4.c. Sun protection	25 (58)	26 (60)	.76

[*]Statistically significant effects indicated in bold.

[†]Refers to responses indicating beliefs or behaviors conducive to preventing influenza illness.

Usability and Utility of the PCHR-Based Program

Of the 123 participants who completed the baseline survey in time to be included in the study, the average number of days to complete the survey was 1.8 days (range 0-25 days) and 1.4 days (range 0-20 days) among intervention and control group participants, respectively. Among the 119 participants who completed at least one biweekly survey, the mean time to completion among intervention subjects (N = 67) was 3.3 days (range 0-24 days) and among control subjects (N = 52), 3.1 days (range 0-15 days). The mean number of completed biweekly surveys was 6.6 (range 1-7) for the intervention group and 6.7 (range 1-7) for the control group. A total of 80% (99/123) of participants completed the exit survey, with mean times to completion of 6.3 days (range 0-27 days) for the intervention group (N = 56) and 7.6 days (range 0-23 days) for the control group (N = 43). Completion rates were 80% (56/70) and 81% (43/53) among the intervention and control groups, respectively.

Among the participants who completed the exit survey, 78% (77/99) rated the PCHR as "extremely" or "very" easy to use and 84% (83/99) indicated survey questions were "extremely" or "very" clear. When asked about specific parts of the messaging system, the aspects deemed most useful by participants were messages with information on prevention of influenza illness, general information on influenza illness, and messages indicating influenza activity in participant's' geographic area. Overall, 73% (72/99) responded that they would be "extremely" or "very" likely to participate again in the use of a PCHR-based health promotion program such as this one. In terms of privacy concerns for providing information electronically, 57% (56/99) were "not at all" or "a little"

concerned, and an additional 25% (25/99) were "moderately" concerned. A total of 62% (61/99) indicated that they would have been willing to provide additional health-related information. Participants found the biweekly surveys to be brief, with 51% (50/99) responding that completion took less than 5 minutes.

Among participants in the intervention group, 54% (30/56) rated the messaging system as "extremely" or "very" useful in providing information about influenza, 13% (7/56) indicated that the messaging system was "extremely" or "very" important in their decision about whether to obtain the influenza vaccine for themselves, and 20% (11/56) responded the same regarding its importance in the immunization of household members.

Discussion

This study evaluated the use of a PCHR-based program for the promotion of positive health behaviors in a workforce population. With a small sample size, the intervention did not demonstrate a significant effect on the majority of the knowledge, beliefs, and behavior elements tested in either the intervention or the control group. However, the study did demonstrate the feasibility of using a PCHR for health promotion in the workplace, with timely responses from participants, high completion rates, and positive feedback from participants regarding the usability and utility of the PCHR-based program.

The small sample size limits interpretation of our results. This was, in part, due to the timing of corporate restructuring at Hewlett Packard, which occurred during the initial recruitment period and resulted in a reduction in eligible participants as well as a decrease in employee interest in a research trial. A post hoc power calculation reveals that given the number of subjects enrolled, a 28% difference in outcome rates between the two arms would have been required to reject the null hypothesis. The majority of the intergroup and intragroup comparisons trended toward a positive effect but did not reach statistical significance, which may be attributable to low power. Another limitation is the short duration of the trial to assess changes in people's beliefs and behaviors surrounding health issues. Sustained education and messaging spanning a second influenza season might strengthen the intervention.

One of the strengths of this type of PCHR-based program is that it could be implemented in most work settings in which employees have Internet access. Prior studies have established the feasibility of Web-based health promotion programs with good enrollment and retention rates [16,17], demonstrating employee acceptance of such interventions. There were very few exclusion criteria in our study, and with a Spanish-language version, this type of program would be accessible to most US employees in diverse work settings and geographic areas. The Web-based format gained high acceptance and gave participants flexibility in deciding when to complete the surveys. Our high completion rates are reflective of the convenience and brevity of the intervention, which made it generally appealing to employees.

Another advantage of this type of program is that it can be tailored to a variety of settings and health issues. Potential settings include clinic populations, student bodies, and members of specific organizations such as smoking cessation groups. Issues ranging from nutrition and weight control to binge drinking, safety belt use, and diabetes management could be targeted [18,19]. Within a given setting, appropriate interventions could also be chosen based on the content of the PCHR and the known health issues of the user.

To our knowledge, this is the first study to examine the use of PCHRs as a tool for health promotion in an employee health program. Although the program did not significantly improve the majority of knowledge, beliefs, and behaviors surrounding influenza prevention, the results are promising enough to suggest benefit in a larger follow-up study over a longer period of time.

Overall, this study provides important evidence for the feasibility and utility of using PCHR-based programs for workplace health promotion. There is a growing movement for employers to offer health promotion services in the workplace [20], and several large companies are in the process of implementing PCHRs for their employees [21]. PCHR-based programs provide a flexible, easily accessible option that can be readily adapted to the specific needs of a workforce population or an individual. Further studies are warranted to explore the use of PCHR-based employee health promotion programs and to identify health issues most suitable to this type of program.

Acknowledgments

This work was supported by Grant 8 R01 CD000286-03 and 5 P01 CD000260-02 from the Centers for Disease Control and Prevention and by National Research Service Awards T32 HD40128-03 and T32 HD40128-04.

We are grateful to the employees at Hewlett Packard who participated in this study. We thank the PING/Indivo development team, Jonathan Abbett, Lucas Jordan, and Nathan Finstein. We are grateful to Donald Goldmann, Dennis Ross-Degnon, and Jonathan Finkelstein for their input on study design. We thank Richard Zippel and Patricia Travers for incorporating the trial into the Hewlett Packard work sites.

Conflicts of Interest

During the formulation, conduct, and analysis there was no conflict of interest. After the data analysis was completed, but during the preparation of the manuscript, Children's Hospital Boston entered into a contract with the nonprofit entity Dossia to support the use of Indivo—an open source, freely available PCHR developed by the Children's Hospital Informatics Program — by the

employees of the Dossia founding companies. The core PCHR software produced under this contract is made freely available as part of the open source code base of Indivo. Some of the authors receive compensation from Children's Hospital Boston to provide advice or technical expertise informing the joint work between Children's Hospital Boston and Dossia, as well as other Indivo deployments. The advice is nonexclusive and unrestricted.

Multimedia Appendix 1

PING screenshots

[PDF file (Adobe PDF), 112 KB - jmir_v10i1e5_app1.pdf]

Multimedia Appendix 2

Enrollment survey (abbreviated)

[PDF file (Adobe PDF), 142 KB - jmir_v10i1e5_app2.pdf]

Multimedia Appendix 3

Biweekly survey

[PDF file (Adobe PDF), 118 KB - jmir_v10i1e5_app3.pdf]

Multimedia Appendix 4

Sample health message

[PDF file (Adobe PDF), 70 KB - jmir_v10i1e5_app4.pdf]

Multimedia Appendix 5

Control group monthly bulletin

[PDF file (Adobe PDF), 70 KB - jmir_v10i1e5_app5.pdf]

References

1. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006 Mar;13(2):121-126 [FREE Full text] [Medline: 16357345] [doi: 10.1197/jamia.M2025]
2. Adams PF, Hendershot GE, Marano MA; Centers for Disease Control and Prevention/National Center for Health Statistics. Current estimates from the National Health Interview Survey, 1996. Vital Health Stat 10 1999 Oct(200):1-203. [Medline: 15782448]
3. Nichol KL, Lind A, Margolis KL, Murdoch M, McFadden R, Hauge M, et al. The effectiveness of vaccination against influenza in healthy, working adults. N Engl J Med 1995 Oct 5;333(14):889-893 [FREE Full text] [Medline: 7666874] [doi: 10.1056/NEJM199510053331401]
4. Smith NM, Bresee JS, Shay DK, Uyeki TM, Cox NJ, Strikas RA; Advisory Committee on Immunization Practices. Prevention and Control of Influenza: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2006 Jul 28;55(RR-10):1-42 [FREE Full text] [Medline: 16874296]
5. Mandl KD, Szolovits P, Kohane IS. Public standards and patients' control: how to keep electronic medical records accessible but private. BMJ 2001 Feb 3;322(7281):283-287 [FREE Full text] [Medline: 11157533] [doi: 10.1136/bmj.322.7281.283]
6. Yasnoff WA, Humphreys BL, Overhage JM, Detmer DE, Brennan PF, Morris RW, et al. A consensus action agenda for achieving the national health information infrastructure. J Am Med Inform Assoc 2004 Jul;11(4):332-338 [FREE Full text] [Medline: 15187075] [doi: 10.1197/jamia.M1616]
7. Morales Rodriguez M, Casper G, Brennan PF. Patient-centered design. The potential of user-centered design in personal health records. J AHIMA 2007 Apr;78(4):44-6; quiz 49-50. [Medline: 17455846]
8. Baker AM, McCarthy B, Gurley VF, Yood MU. Influenza immunization in a managed care organization. J Gen Intern Med 1998 Jul;13(7):469-475 [FREE Full text] [Medline: 9686713] [doi: 10.1046/j.1525-1497.1998.00136.x]
9. ; Indivo Health: Personally Controlled Health Record. URL: http://www.indivohealth.org/ [accessed 2008 Feb 24] [WebCite Cache ID 5SmZMFzqQ]
10. Riva A, Mandl KD, Oh DH, Nigrin DJ, Butte A, Szolovits P, et al. The personal internetworked notary and guardian. Int J Med Inform 2001 Jun;62(1):27-40. [Medline: 11340004] [doi: 10.1016/S1386-5056(00)00136-2]
11. Simons WW, Mandl KD, Kohane IS. The PING personally controlled electronic medical record system: technical architecture. J Am Med Inform Assoc 2005 Jan;12(1):47-54 [FREE Full text] [Medline: 15492031] [doi: 10.1197/jamia.M1592]

12. Mandl KD, Simons WW, Crawford WCR, Abbett JM. Indivo: a personally controlled health record for health information exchange and communication. BMC Med Inform Decis Mak 2007;7(1):25 [FREE Full text] [Medline: 17850667] [doi: 10.1186/1472-6947-7-25]
13. Harper SA, Fukuda K, Uyeki TM, Cox NJ, Bridges CB; Advisory Committee on Immunization Practices (ACIP), Centers for Disease Control and Prevention (CDC). Prevention and control of influenza. Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2005 Jul 29;54(RR-8):1-40. [Medline: 16086456]
14. 121 Cities Mortality Reporting System. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/EPO/DPHSI/121hist.htm [accessed 2007 Oct 22] [WebCite Cache ID 5SmZUempz]
15. ; Healthy People 2010. US Department of Health and Human Services. URL: http://www.healthypeople.gov/document/pdf/uih/2010uih.pdf [accessed 2008 Feb 24] [WebCite Cache ID 5WGPOpaXD]
16. Franklin PD, Rosenbaum PF, Carey MP, Roizen MF. Using sequential e-mail messages to promote health behaviors: evidence of feasibility and reach in a worksite sample. J Med Internet Res 2006;8(1):e3 [FREE Full text] [Medline: 16585028] [doi: 10.2196/jmir.8.1.e3]
17. Cook RF, Billings DW, Hersch RK, Back AS, Hendrickson A. A field test of a web-based workplace health promotion program to improve dietary practices, reduce stress, and increase physical activity: randomized controlled trial. J Med Internet Res 2007;9(2):e17 [FREE Full text] [Medline: 17581811] [doi: 10.2196/jmir.9.2.e17]
18. Linke S, Murray E, Butler C, Wallace P. Internet-based interactive health intervention for the promotion of sensible drinking: patterns of use and potential impact on members of the general public. J Med Internet Res 2007;9(2):e10 [FREE Full text] [Medline: 17513281] [doi: 10.2196/jmir.9.2.e10]
19. Saul JE, Schillo BA, Evered S, Luxenberg MG, Kavanaugh A, Cobb N, et al. Impact of a statewide Internet-based tobacco cessation intervention. J Med Internet Res 2007;9(3):e28 [FREE Full text] [Medline: 17942390] [doi: 10.2196/jmir.9.4.e28]
20. Okie S. The employer as health coach. N Engl J Med 2007 Oct 11;357(15):1465-1469. [Medline: 17928592] [doi: 10.1056/NEJMp078152]
21. Dossia. URL: http://www.dossia.org [accessed 2008 Feb 24] [WebCite Cache ID 5SmZfkosI]

Abbreviations

ACIP: Advisory Committee on Immunization Practices
CDC: Centers for Disease Control and Prevention
PCHR: personally controlled health record

submitted 22.10.07; peer-reviewed by O Nytro; comments to author 14.11.07; revised version received 03.02.08; accepted 09.02.08; published 14.03.08.

Please cite as:
Bourgeois FT, Simons WW, Olson K, Brownstein JS, Mandl KD
Evaluation of Influenza Prevention in the Workplace Using a Personally Controlled Health Record: Randomized Controlled Trial
J Med Internet Res 2008;10(1):e5
URL: http://www.jmir.org/2008/1/e5/
doi:10.2196/jmir.984
PMID:18343794

Original Paper

Long-Term Patterns of Online Evidence Retrieval Use in General Practice: A 12-Month Study

Farah Magrabi[1], BE, PhD; Johanna I Westbrook[2], MHA, PhD, FACHI, FACMI; Michael R Kidd[3], MB, BS, MD, FRACGP, FACHI; Richard O Day[4,5], AM, MB, BS, MD, FRACP; Enrico Coiera[1], MB, BS, PhD, FACMI

[1]Centre for Health Informatics, University of New South Wales, Sydney, Australia

[2]Health Informatics Research & Evaluation Unit, Faculty of Health Sciences, University of Sydney, Sydney, Australia

[3]Discipline of General Practice, University of Sydney, Sydney, Australia

[4]Physiology & Pharmacology, School of Medical Sciences, Faculty of Medicine, University New South Wales, Sydney, Australia

[5]Clinical Pharmacology, Therapeutics Centre, St Vincent's Hospital, Sydney, Australia

Corresponding Author:
Farah Magrabi, BE, PhD
Centre for Health Informatics
University of New South Wales
Sydney 2052
Australia
Phone: +61 2 9385 9009
Fax: +61 2 9385 9006
Email: f.magrabi@unsw.edu.au

Abstract

Background: Provision of online evidence at the point of care is one strategy that could provide clinicians with easy access to up-to-date evidence in clinical settings in order to support evidence-based decision making.

Objective: The aim was to determine long-term use of an online evidence system in routine clinical practice.

Methods: This was a prospective cohort study. 59 clinicians who had a computer with Internet access in their consulting room participated in a 12-month trial of Quick Clinical, an online evidence system specifically designed around the needs of general practitioners (GPs). Patterns of use were determined by examination of computer logs and survey analysis.

Results: On average, 9.9 searches were conducted by each GP in the first 2 months of the study. After this, usage dropped to 4.4 searches per GP in the third month and then levelled off to between 0.4 and 2.6 searches per GP per month. The majority of searches (79.2%, 2013/2543) were conducted during practice hours (between 9 am and 5 pm) and on weekdays (90.7%, 2315/2543). The most frequent searches related to diagnosis (33.6%, 821/2291) and treatment (34.5%, 844/2291).

Conclusion: GPs will use an online evidence retrieval system in routine practice; however, usage rates drop significantly after initial introduction of the system. Long-term studies are required to determine the extent to which GPs will integrate the use of such technologies into their everyday clinical practice and how this will affect the satisfaction and health outcomes of their patients.

(J Med Internet Res 2008;10(1):e6) doi:10.2196/jmir.974

KEYWORDS

Clinical informatics; information retrieval; evidence-based medicine; family practice; evaluation studies; Internet

Introduction

Good quality online evidence retrieval systems should provide clinicians with convenient access to up-to-date, reliable, and pertinent information at the point of care. While the potential of online evidence systems in providing information to answer clinical questions has been demonstrated in controlled laboratory settings [1], their impact on clinicians' decision-making behavior is dependent on uptake and sustained use in an everyday clinical setting. Of the few investigations of online evidence use in routine clinical work, the majority have measured usage by clinicians a few weeks following provision of the system [2,3]. There are few long-term assessments beyond the initial period of introduction, during which the perceived novelty of the intervention is likely to affect patterns of use.

We sought to measure the long-term use of an online evidence retrieval system, Quick Clinical (QC), in routine general practice settings. In a previous 4-week study conducted from October to November 2002, 193 general practitioners (GPs) used the QC online evidence system to perform an average of 8.7 searches per month [4]. The majority of these searches (81.1%) were conducted from consulting rooms during office hours. The most frequent searches related to diagnosis (37.3%) and treatment (32.1%). Search topics included a broad spectrum of diseases, including common conditions such as asthma, diabetes, and hypertension. In this paper we present the results of a 12-month trial of QC in general practice.

Methods

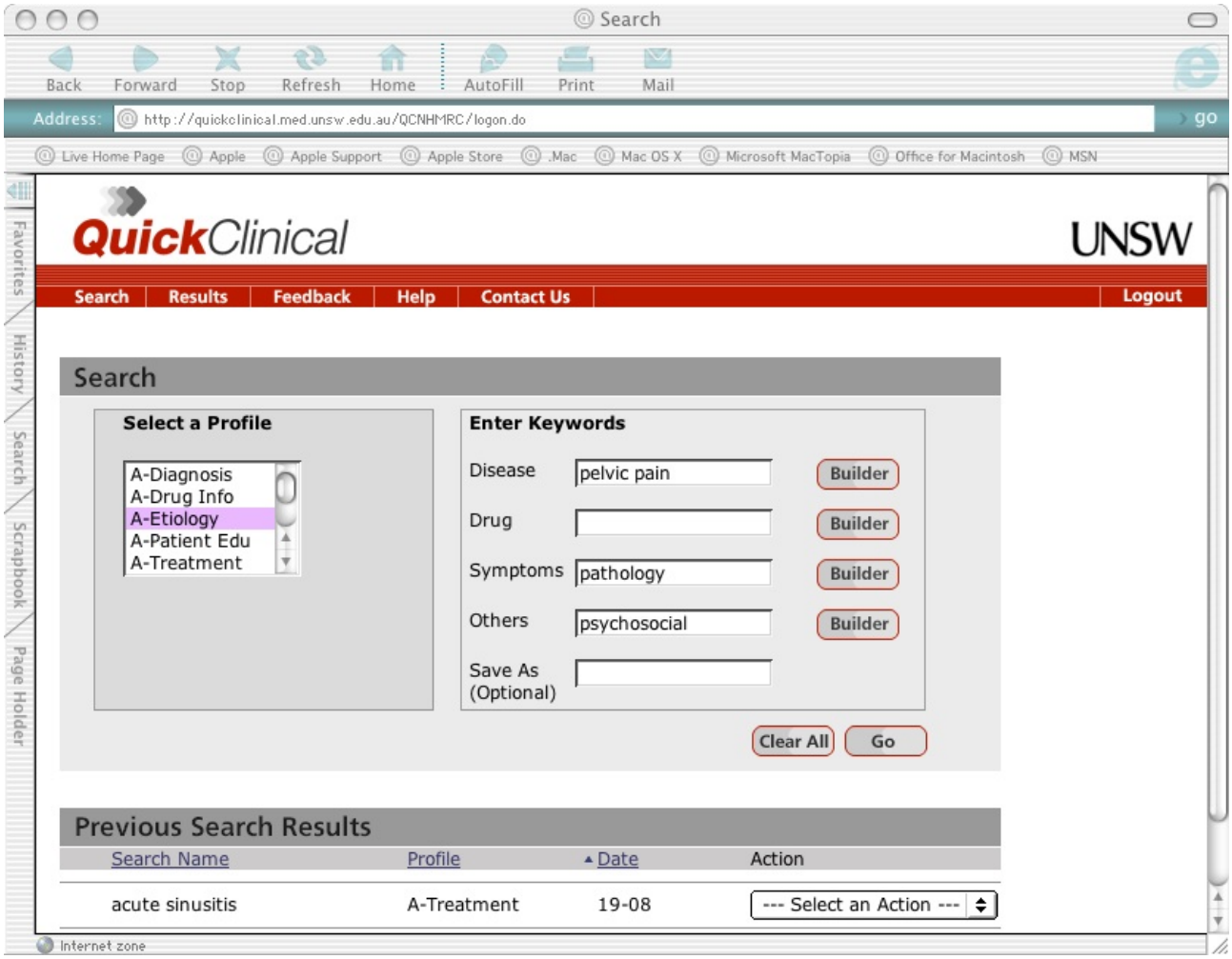
Setting and Participants

A total of 59 GPs from across Australia participated in the trial. Clinicians who had a computer with Internet access in their consulting room were recruited via a call for volunteers advertised in journals, newsletters, and a clinician listserv.

Quick Clinical

QC is based on the generic use of search filters explicitly designed to meet the information needs of specific user groups. The filters can be customized to meet the varying needs of different groups [5]. Search filters were adjusted for this study to provide five "profiles" specifically designed for GPs: disease etiology, diagnosis, treatment, prescribing, and patient education. Users first select a search filter or profile that matches their question type (eg, diagnosis, treatment) and then enter keywords that more specifically describe their query. Up to four types of keywords can be used in association with a given profile: disease, drug, symptoms, other. For example, a clinician who encounters a 32-year-old woman with a fourth presentation of pelvic pain in the last 6 months but whose physical examination, ultrasound studies, and swabs for infection are all negative, may have a question regarding the social, psychological, as well as biological causes of pelvic pain. The clinician could select the "etiology" profile and enter "pelvic pain," "pathology," and "psychosocial" as keywords (Figure 1). The search filters retrieve evidence from information resources selected for local relevance, including PubMed, MIMS (a pharmaceutical database), Therapeutic Guidelines, Merck Manual, and HealthInsite (a government-funded health database for consumers [6]). Users can also search each of these resources individually.

Figure 1. User interface of QC



Procedures

Clinicians were asked to use QC in their practice from May 2005 to April 2006. QC was available via a standard Web browser interface (eg, Firefox, Microsoft Internet Explorer). Each participant obtained a personal username and password to access QC and completed an online tutorial on how to use the system. A help manual was also available online. All participants were asked to complete an online survey about their computer use during consultations, and demographic information was also sought at the beginning of the study. We did not send out any reminders or prompt participants to keep using QC. Frequency and purpose of system use were determined from automatically generated computer logs used to record details of each search, including the search filter chosen, keywords entered, data sources accessed, and the date, time, and duration of the searches.

The study was recognized by the Royal Australian College of General Practitioners (RACGP) for its continuing medical education (CME) program. Education points were not directly linked to the number of searches performed, but to trial completion. Ethics approval for the protocol was received from the ethics committees of the University of New South Wales, University of Sydney, and RACGP.

Analysis

Statistical analysis of data from the computer logs and online survey was undertaken using SPSS (Statistical Package for the Social Sciences) v11.0 (SPSS Inc, Chicago, IL, USA). Descriptive statistics were used to examine patterns of use and responses to the online survey. Comparisons between groups were made using Student *t* test and chi-square analyses.

Results

Participants

A total of 140 GPs expressed interest in using QC in their practice. Of these, only 59 (42%) completed the online registration and tutorial enabling them to use the system. The majority of participants were male (71%, 42/59), aged 35-54 years, and 71% (42/59) obtained their primary medical qualification in Australia. Most GPs worked in a group practice or medical center, and 56% (33/59) were fellows of the RACGP (Table 1). The majority (83%, 42/59) worked in an accredited practice, and 78% (46/59) had 11 or more years experience in primary care. On average, participants worked 34.16 hours per week in direct patient care (SD = 10.48) and consulted with 4.07 patients per hour (SD = 0.69).

Table 1. Demographics of study GPs in comparison to the general population of Australian GPs in 2005/2006

Characteristic	%	
	Study GPs (N = 59)	Australian GPs* (N = 953)
Gender (χ^2_1 = 0.3, *P* = .60)		
Male	71	67.9
Female	29	32.1
Age (χ^2_3 = 3.8, *P* = .28)		
< 35 years	8	8.9
35-44 years	32	25.5
45-54 years	37	31.8
55+ years	22	33.8
Country of graduation (χ^2_1 = 0.04, *P* = .83)		
Australia	71	69.9
Overseas	29	30.1
Fellow of RACGP (χ^2_1 = 4.2, *P* = .04)	56	42.3
Practice type (χ^2_1 = 1.2, *P* = .28)		
Group or medical center	83	87.8
Solo	17	12.2
Computer use during consultations (χ^2_4 = 1.3, *P* = .87)†		
Prescribing	100	89.5
Medical records	93	76.4
Internet	80	72.9
Other administrative purposes (eg, appointments)	78	79.8
Email	75	72.9
Patient education	95	62.9‡
Online evidence	81	16.6‡

*Data are for Australian GPs in 2005/2006 [7].

†N = 880 for computer use data among Australian GPs [7].

‡N = 1061 for patient education and online evidence data among Australian GPs [8].

All participants reported having a computer on their desk where they saw patients. All but one used their computer during consultations, and 88% (52/59) indicated having "good" to "excellent" computer skills. Computers were used for a range of practice functions, including prescribing, medical records, practice administration, Internet, email, patient education, and online evidence. Of those who knew their Internet connection type, 89% (47/53) reported having access via a broadband connection.

Patterns of QC Use

In total, participants conducted 2543 searches over the 12-month period (May 2005 to April 2006). The total number of searches conducted by each participant ranged from 1 to 240 over the trial ($mean_{59}$ = 39.14, SD = 45.29; $median_{59}$ = 23); 9 participants did not use QC after the first 2 months of the study ($mean_{50}$ = 38.28, SD = 38.80; $median_{50}$ = 28). Relatively higher rates of use were recorded in the initial 2 months of the study (Figure 2). On average, 9.1 to 10.8 searches were conducted by each GP during this period. After this, the usage rate dropped to 4.4 searches per GP in the third month and then levelled off to between 0.4 and 2.6 searches per GP per month. There was significant variation in individual use of the system (Figure 3). We compared the group of participants who used QC for less than 10 searches (36%, 21/59), the "low" use group, with those who used the resource 50 or more times (29%, 17/59), the "high" use group. There was no difference in the makeup of the high and low use groups by gender, years of general practice experience, place of graduation, practice type, RACGP fellowship status, or information-seeking behavior (Table 2).

However, the low use group had a significant number of participants aged 45 years and older (χ^2_1= 4.8, *P* = .03).

Figure 2. Average number of monthly QC searches over 12-month study period (N = 2543 searches)

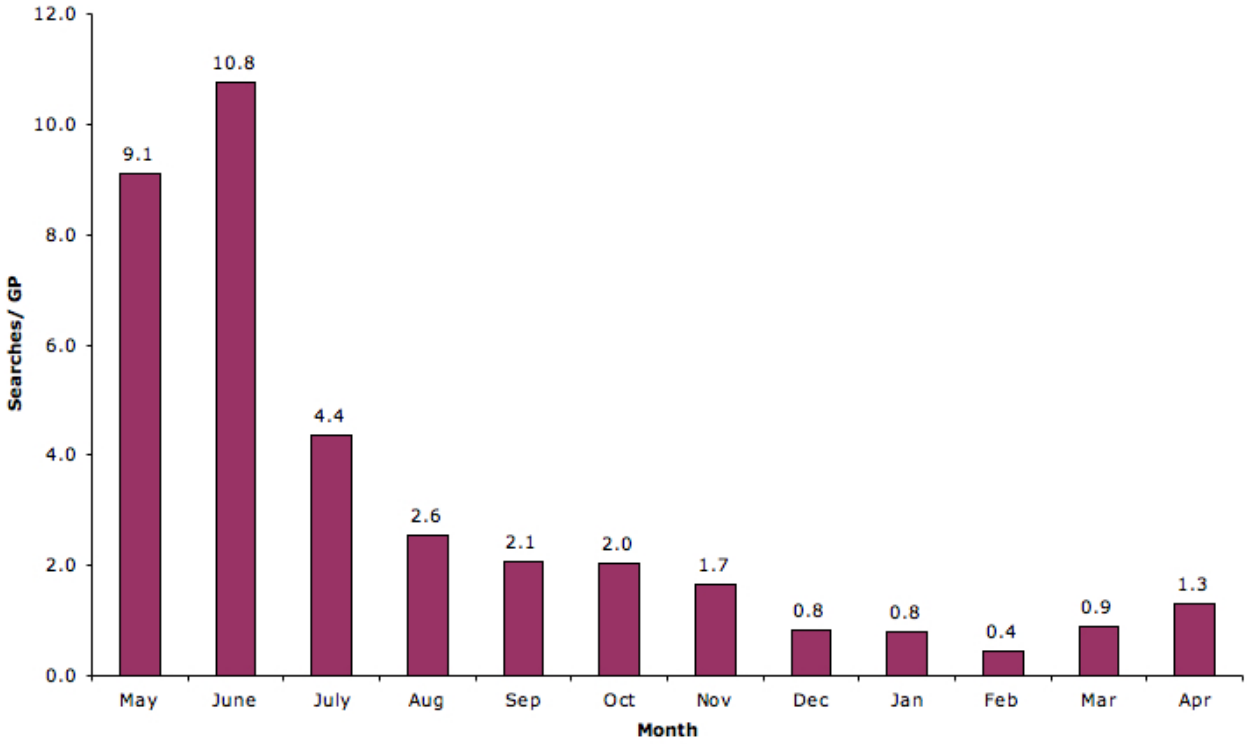


Figure 3. Percentage of GPs conducting QC searches over the 12-month study period, by number of searches (N = 59 GPs)

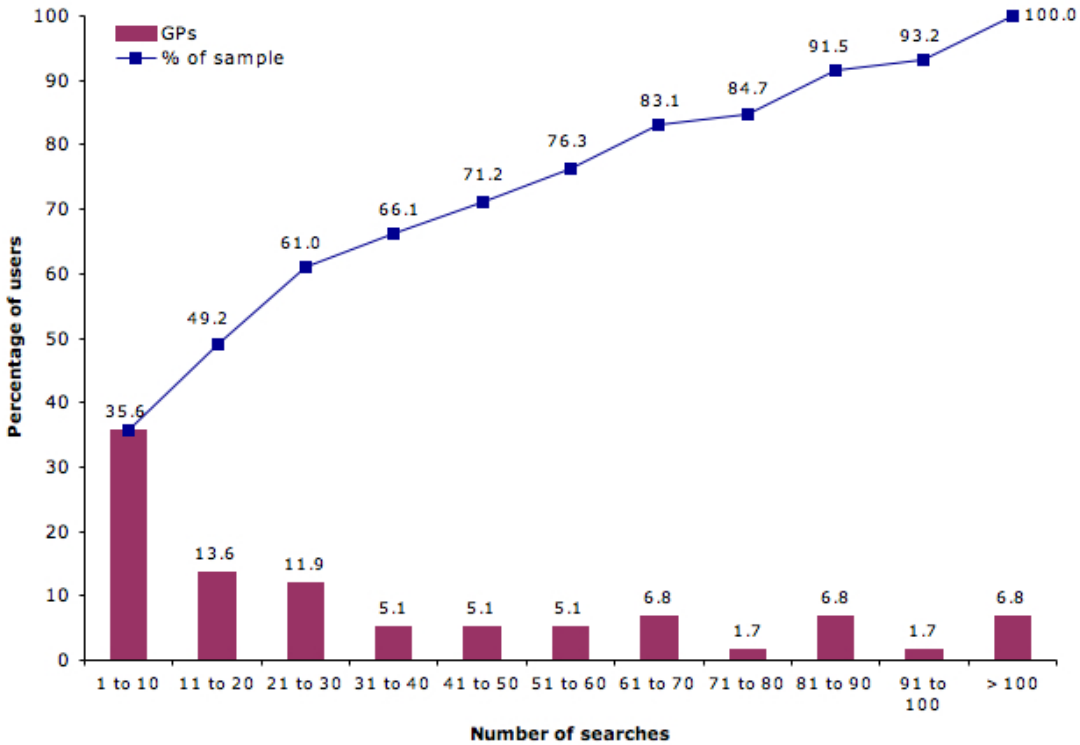


Table 2. Comparison of QC high and low use groups

Characteristic	% (No.)	
	Low Use* (N = 21)	High Use† (N = 17)
Gender (χ^2_1 = 0.6, *P* = .44)		
Female	24 (5)	35 (6)
Male	76 (16)	65 (11)
Age (χ^2_1 = 4.8, *P* = .03)		
< 45 years	19 (4)	53 (9)
45+ years	81 (17)	47 (8)
Country of graduation (χ^2_1 = 0.2, *P* = .64)		
Australia	57 (12)	65 (11)
Overseas	43 (9)	35 (6)
Experience in general practice (χ^2_1 = 1.4, *P* = .24)		
≤ 10 years	10 (2)	24 (4)
11+ years	90 (19)	76 (13)
Practice type (χ^2_1 = 0.2, *P* = .70)		
Group or medical center	76 (16)	71 (12)
Solo	24 (5)	29 (5)
Fellow of RACGP (χ^2_1 = 0.1, *P* = .80)		
Yes	43 (9)	47 (8)
No	57 (12)	53 (9)
Search for information during consultations (χ^2_1 = 0.4, *P* = .54)		
Yes	81 (17)	88 (15)
No	19 (4)	12 (2)

*Low use is 1-10 searches.

†High use is ≥ 50 searches.

QC use varied throughout the day. The system was mostly used during practice hours, peaking in the morning and afternoon sessions; 79% (2013/2543) of the searches were conducted between 9 am and 5 pm (Figure 4). The use of the system also varied over the work week, peaking on Wednesday; 91% (2315/2543) of the searches were conducted between Monday and Friday (Figure 5). Thus, some use also occurred outside work hours.

Figure 4. QC use by time of day (12-month N = 2543; 4-week N = 1257 searches)

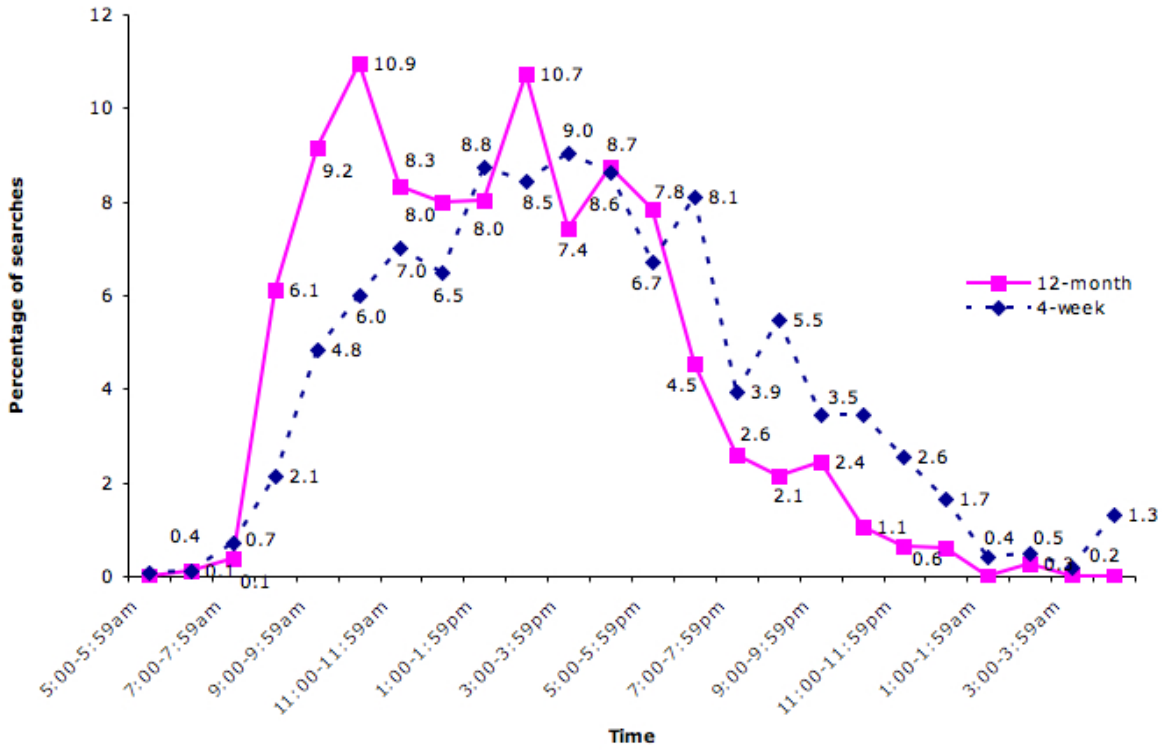
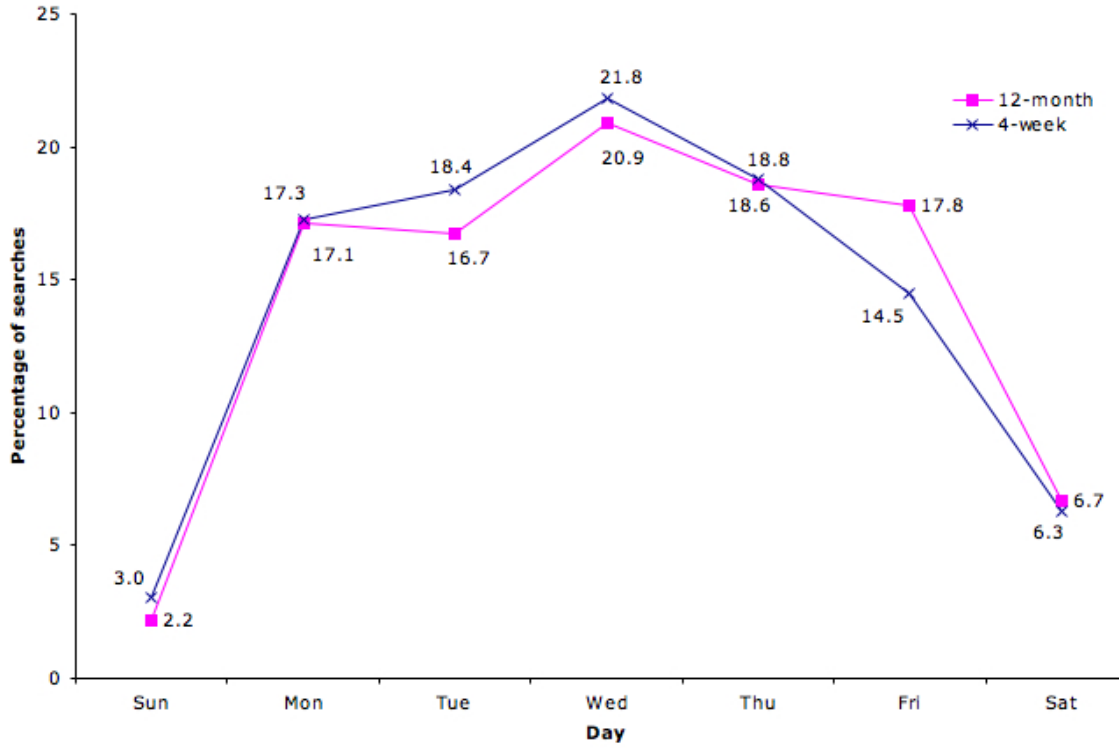


Figure 5. QC use by day of the week (12-month N = 2543; 4-week N = 1293 searches)



Search Patterns

A large proportion of searches (90%, 2291/2543) were undertaken using a QC profile. Of these, 33.6% (821/2291) related to questions about diagnosis, 34.5% (844/2291) to treatment, and 13.8% (337/2291) to patient education. Almost 5% (117/2291) were related to prescribing and 6.0% (147/2291) to disease etiology. Disease-specific keywords were used to describe clinical questions in a significant proportion of the searches (72.9%, 1854/2543). In comparison, few searches utilized keywords related to drugs (6.5%, 165/2543) or symptoms (8.8%, 224/2543). The fourth keyword type, "other,"

was utilized in 19.3% (491/2543) of the searches. The 10 most frequently used entries for each keyword category are listed in Table 3.

Table 3. Top 10 keywords used to describe clinical questions (N = 2543 searches)

Disease (n = 1854)	Drug (n = 165)	Symptoms (n = 224)	Other (n = 491)
asthma	self-examination	pain	prevention
maculopathy	Implanon	abdominal	cholecystectomy
dermatitis	vitamins	itch	breast self-examination
ovarian cancer	folic acid	nocturia	pregnancy
gout	isotretinoin	upper ab	females
cholecystectomy	mirtazapine	lumps	diet
acne	flucloxacillin	pain	inheritance
cholecystitis	prednisolone	gallstone	surgery
sudden infant death syndrome	dietary supplement	hematur	dietary sources
breast cancer	saw palmetto	lump	child

Discussion

Main Findings and Implications

This is the first study to directly measure individual GPs' long-term patterns of use of an online evidence facility. We found that QC was used mostly during weekday practice hours. On average, each clinician conducted 0.7 searches per month over the trial. The majority of the searches related to questions about diagnosis and treatment. These findings indicate that the QC model fits into general practice and that GPs will use online evidence past a typical 1- or 2-month trial.

On average, each clinician used QC for one search every 2 months. Although the use of electronic resources is reported to be growing steadily, a recent review confirms that GPs still prefer to consult their colleagues and textbooks over electronic resources to answer their clinical questions [9]. In a study of clinicians' actual information-seeking behavior, Ely et al [10] found online sources to be the third most frequently used resource after textbooks and humans. Depending on practice variables and methods used to measure the frequency of clinical questions, it is estimated that GPs generate up to two questions per consultation [9]. Given that clinicians typically pursue only a small proportion of their clinical questions, the current long-term frequencies of online evidence use are likely to be low.

After an initial surge in the use of QC during the first 2 months of the study, utilization by each GP levelled to 0.6 searches per month on average. This could be attributed to the natural tendency for usage rates to drop as the novelty of the system wears off. Lack of easy access may also account for low rates of use as participants were required to log on to QC each time they used it. The use of other search systems may have been a confounding variable. At the beginning of the study, 81% (48/59) of participants reported that they used online evidence during consultations, indicating that online information seeking was not a novel practice to them and that QC may have been used alongside a range of other online resources they regularly consulted. As this was not a controlled study, other variables might have affected use independent of QC.

A significant proportion of the low use group was 45 years or older (Table 2); it is possible that this group may have been less comfortable with online searching despite their reports of having good to excellent computer skills. It is also likely that changes in practice conditions may have impacted QC use. Only 59 of the 140 GPs who initially registered completed the study. Relocations and changes in practice conditions may have impacted participation in this study. However, in a follow-up of a previous study of QC in which we specifically examined factors associated with integration of online evidence into clinical practice, we found that levels of use could only be directly linked to clinicians' experiences of improvement in patient care as a result of using QC [11].

There are few studies of the type and quality of online resources used by clinicians. In a laboratory study that allowed participants to choose their own electronic resources to answer simulated clinical questions, investigators found that despite the availability of high-quality resources such as Clinical Evidence and Cochrane, Google and other Internet sites were used at the same rate as Medline (22.6%) and accounted for the second most frequently used resources after UpToDate (65.9%), a resource presenting concise summaries of clinical evidence [12]. Increasing use of general purpose search engines to retrieve online information of variable quality is likely to impact the quality and safety of clinical decisions, a trend worthy of monitoring.

Comparison With Existing Literature

Little comparative data are available as there are few studies of online evidence retrieval use in general practice. Clinicians' use of QC beyond the initial 2 months is comparable to studies of Medline use in hospital and ambulatory settings. Thus, eight out of 10 studies reviewed by Hersh [2] reported utilization rates ranging from 0.3 to 3.5 searches per person-month over 2 to 36 months. Our study data are comparable to these data. The two outliers are both short-term studies: Collen and Flagle's 2.7-month study reported 6.7 searches per person-month [13],

and Osheroff and Bankowitz's 2-week study reported 12.5 searches per person-month [14]. In contrast with clinicians' self-reports of online information seeking, where 45% of searching was reported to occur outside practice hours [15], we found that QC was largely used during practice hours (79% of searches were conducted between 9 am and 5 pm).

While use of QC went down after 4 weeks, the overall pattern remained similar in terms of days and times when searches were undertaken. General patterns of QC use observed in the current study are consistent with a 2002 trial of the system over 4 weeks [4]. We found no difference in the overall utilization pattern by time of day (see Figure 4) or day of the week (χ^2_6= 9.7, P = .14, see Figure 5). However, there was a significant difference in QC profile use; while the proportion of diagnosis and etiology questions was similar, a larger proportion of questions in the 12-month study related to diagnosis and patient education, and fewer related to prescribing (14.9%, χ^2_4= 35.7, P < .001). As in the previous 4-week trial, there was considerable variation in use of QC among individual clinicians. While short-term trials are adequate for predicting broad patterns of online evidence use, long-term studies are still necessary to measure the overall uptake and integration into clinical practice.

Limitations of This Study

The participants were a self-selected cohort who volunteered to participate in the study. In the pre-trial survey, eight out of 10 GPs within this group (81%, 48/59) reported using online evidence during consultations and may therefore have been predisposed to using QC in their practice compared to the general population (17%, see Table 1). The majority of participants were new to QC; however, some (29%, 17/59) reported using the system in the previous 2002 study. When compared to the general population of Australian GPs, there was no difference in gender, age, or place of graduation. Participants' overall computer use was found to be representative of that in the general population of GPs within Australia. Though education points were linked to trial completion and not the number of searches performed, recognition of this study as a CME activity is likely to have resulted in a significantly higher proportion of RACGP fellows within our sample. On the whole, the demographics of our cohort were generally comparable to the general population of Australian GPs.

Conclusion

This study measured GPs' individual use of an online evidence retrieval system over a 12-month period. Clinicians used the system in routine care to answer questions mostly about diagnosis and treatment. Usage rates dropped significantly after initial introduction of the system. While short-term trials are adequate for measuring broad patterns of online evidence use, overall uptake and integration into clinical practice require long-term studies.

Acknowledgments

We wish to thank the clinicians who gave their time to take part in this study and Mr. K Nguyen and Mr. V Vickland for maintaining the Quick Clinical evidence retrieval system. This research was partially funded by an Australian National Health and Medical Research Council (NHMRC) Project Grant (300435). Financial support to develop Quick Clinical was provided by an Australian Research Council SPIRT grant (C00107730) and an NHMRC Development Grant (300591). JW is supported by an NHMRC Fellowship.

Conflicts of Interest

The University of New South Wales and Enrico Coiera could benefit from the commercial exploitation of the Quick Clinical search engine or its technologies.

References

1. Westbrook JI, Coiera EW, Gosling AS. Do online information retrieval systems help experienced clinicians answer clinical questions? J Am Med Inform Assoc 2005;12(3):315-321. [Medline: 15684126] [doi: 10.1197/jamia.M1717]
2. Hersh W. Information Retrieval. A Health and Biomedical Perspective. 2nd edition. New York: Springer Verlag; 2003:528.
3. Pluye P, Grad RM, Dunikowski LG, Stephenson R. Impact of clinical information-retrieval technology on physicians: a literature review of quantitative, qualitative and mixed methods studies. Int J Med Inform 2005 Sep;74(9):745-768. [Medline: 15996515] [doi: 10.1016/j.ijmedinf.2005.05.004]
4. Magrabi F, Coiera EW, Westbrook JI, Gosling AS, Vickland V. General practitioners' use of online evidence during consultations. Int J Med Inform 2005 Jan;74(1):1-12. [Medline: 15626631] [doi: 10.1016/j.ijmedinf.2004.10.003]
5. Coiera E, Walther M, Nguyen K, Lovell NH. Architecture for knowledge-based and federated search of online clinical evidence. J Med Internet Res 2005;7(5):e52 [FREE Full text] [Medline: 16403716] [doi: 10.2196/jmir.7.5.e52]
6. HealthInsite. URL: http://www.healthinsite.gov.au/ [accessed 2008 Feb 19] [WebCite Cache ID 5UntPdmdc]
7. Britt H, Miller GC, Knox S, Charles J, Pan Y, Henderson J, et al. General Practice Activity in Australia 2004-05. Canberra, Australia: The University of Sydney and the Australian Institute of Health and Welfare; 2005.
8. McInnes DK, Saltman DC, Kidd MR. General practitioners' use of computers for prescribing and electronic health records: results from a national survey. Med J Aust 2006 Jul 17;185(2):88-91 [FREE Full text] [Medline: 16842064]
9. Coumou HCH, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. J Med Libr Assoc 2006 Jan;94(1):55-60. [Medline: 16404470]

10. Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. Answering physicians' clinical questions: obstacles and potential solutions. J Am Med Inform Assoc 2005;12(2):217-224 [FREE Full text] [Medline: 15561792] [doi: 10.1197/jamia.M1608]
11. Magrabi F, Westbrook JI, Coiera EW. What factors are associated with the integration of evidence retrieval technology into routine general practice settings? Int J Med Inform 2007 Oct;76(10):701-709. [Medline: 16893677] [doi: 10.1016/j.ijmedinf.2006.06.009]
12. McKibbon KA, Fridsma DB. Effectiveness of clinician-selected electronic information resources for answering primary care physicians' information needs. J Am Med Inform Assoc 2006;13(6):653-659 [FREE Full text] [Medline: 16929042] [doi: 10.1197/jamia.M2087]
13. Collen MF, Flagle CD. Full-text medical literature retrieval by computer. A pilot test. JAMA 1985 Nov 15;254(19):2768-2774. [Medline: 3903234] [doi: 10.1001/jama.254.19.2768]
14. Osheroff JA, Bankowitz RA. Physicians' use of computer software in answering clinical questions. Bull Med Libr Assoc 1993 Jan;81(1):11-19 [FREE Full text] [Medline: 8428184]
15. Bennett NL, Casebeer LL, Zheng S, Kristofco R. Information-seeking behaviors and reflective practice. J Contin Educ Health Prof 2006;26(2):120-127. [Medline: 16802313] [doi: 10.1002/chp.60]

Abbreviations

CME: continuing medical education
GP: general practitioner
QC: Quick Clinical
RACGP: Royal Australian College of General Practitioners

submitted 13.09.07; peer-reviewed by D Little; comments to author 10.01.08; revised version received 07.02.08; accepted 11.02.08; published 19.03.08.

Please cite as:
Magrabi F, Westbrook JI, Kidd MR, Day RO, Coiera E
Long-Term Patterns of Online Evidence Retrieval Use in General Practice: A 12-Month Study
J Med Internet Res 2008;10(1):e6
URL: http://www.jmir.org/2008/1/e6/
doi:10.2196/jmir.974
PMID:18353750

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org