Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 9 (2007), Issue 2 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Viewpoints

Original Papers

Viewpoint

Improving Information Technology Adoption and Implementation Through the Identification of Appropriate Benefits: Creating IMPROVE-IT

Kevin J Leonard[1], BComm, MBA, PhD, CMA; Dean F Sittig[2], BS, MS, PhD

[1]Department of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[2]Applied Research in Medical Informatics, Northwest Permanente and Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA

Corresponding Author:
Kevin J Leonard, BComm, MBA, PhD, CMA
Associate Professor, Department of Health Policy, Management and Evaluation
Faculty of Medicine, University of Toronto
12 Queens Park Crescent West, 2nd Floor, McMurrich Building
Toronto, ON M5S 1A8
Canada
Phone: +1 416 978 8364
Fax: +1 416 978 7350
Email: k.leonard@utoronto.ca

Abstract

This paper describes the objectives of a collaborative initiative that attempts to provide the evidence that increased information technology (IT) capabilities, availability, and use lead directly to improved clinical quality, safety, and effectiveness within the inpatient hospital setting. This collaborative network has defined specific measurement indicators in an attempt to examine the existence, timing, and level of improvements in health outcomes that can be derived from IT investment. These indicators are in three areas: (1) IT costs (which includes both initial and ongoing investment), (2) IT infusion (ie, system availability, adoption, and deployment), and (3) health performance (eg, clinical efficacy, efficiency, quality, and effectiveness). Herein, we outline the theoretical framework, the methodology employed to create the metrics, and the benefits that can be obtained.

(J Med Internet Res 2007;9(2):e9) doi:10.2196/jmir.9.2.e9

KEYWORDS

Information technology; IT cost; IT investment; Timing of benefits; Efficiencies; Cost-effectiveness; Health outcomes

Introduction

A critical question is now facing health care: Does spending on information technology (IT) lead to greater system availability, increased clinician use, improved decision making, and better health outcomes? It is believed that one of the reasons health care systems have not widely adopted IT is that the benefits that emanate from investment in IT are poorly, if at all, defined [1-3]. In an attempt to address this need, this research initiative has developed several measures that link clinical system availability, use, and cost to clinical impact over a wide range of health care scenarios. Overall, it is the goal of IMPROVE-IT (indices measuring performance relating outcomes, value and expenditure through information technology) to demonstrate the relationship between IT and better health outcomes. Ultimately, the IMPROVE-IT project is attempting to provide a basis for the creation and dissemination of the evidence that increased IT capabilities, availability, and use lead directly to improved clinical quality, safety, and effectiveness, focusing primarily within the inpatient hospital setting [4].

The research literature has discussed the need for measuring the value associated with IT [5,6]. In order to accomplish this, there is a need to develop better methods for tracking IT spending, system availability, and utilization. Recent studies have attempted to estimate the business value generated from IT investment in health care in specific areas, but they have not recommended any method for measuring a broader (eg, hospital-wide) effect or for dealing with the problems of partial implementation [7-11]. In addition, studies are now underway to establish long-awaited quality benchmarks and performance measures for the health system across a number of perspectives, such as physician adherence to best practices [12], availability of systems [13], statistical modeling to predict efficiency [14], and overall scorecard development [15]. However, there has

yet to be a detailed study examining the relationship between IT investment and improvement in financial efficiencies *and* health performance outcomes, where the results and benefits are observed.

It is hypothesized that IT spending provides an environment for a new and comprehensive level of care to exist. That is, without new technology and better information, clinicians would not be able to deliver the effective care that they can when these types of investment are made. IT can provide an opportunity to assess trends that formerly have taken much longer to identify. Improved information access can lead to rapid decision making relating to that information. Often, a clinical decision support tool is a component of the new system. Decisions aided by this support system may improve the operation of the organization: actions can now be taken sooner than they were historically, if they were taken at all. Finally, IT can be used to evaluate its own effectiveness by providing information on the improvement across a wide range of indicators. In short, better information can lead to better care, as demonstrated by improved health outcomes, such as creating the ability to diagnose patients more accurately, as well as sooner; to comply with patients' wishes; to reduce the number or severity of errors; and to support care delivery through better access to information.

Although capturing and documenting IT investment, utilization, and outcomes may appear, on the surface, to be straightforward, there are many disparate factors and complexities that make this process extremely difficult. For example, while the concept of an electronic health record (EHR) seems clear, we believe that to actually create the infrastructure required to support EHRs, one must include many more aspects of clinical information and communications technologies. In an effort to simplify this process to some degree, in this paper we will use the term clinical information systems (CISs) to refer to this larger conglomeration of clinical information and communications technology. Secondly, before one can begin to manage such an effort, one must have the means (or availability) to measure that progress. Measuring the extent to which these systems have been deployed and are being used are the first two critical steps in measuring the overall impact these systems will ultimately have on the quality of care received by all patients.

As a result, and borrowing several concepts from conventional quality measurement efforts, it is clear that we must be able to measure aspects of the structure, process, and outcomes that make up these CISs. These concepts translate into measurements of health information management technology availability, use, and effectiveness at many different levels. For example, we believe that we must make these measurements at the single physician level, the clinic level, the hospital level, the entire health care organization level, and even the local, regional, and national levels. In addition, all of these measurements need to be made from the multiple viewpoints of the key users of these systems, namely, patients, clinicians, and those involved in population health activities (eg, public health departments). As in any large-scale measurement and evaluation effort, designing and validating the measures will be one of the most important and difficult challenges to overcome. In the immediate term, the objective is to entrench the philosophy of measurement through the selection of a "pilot" group of indicators being reported on in a hospital setting.

Going forward, contributing hospitals will be asked to provide measures on their hospital's performance each quarter to a secured website. In exchange for this commitment, hospitals will be provided access to the secured website and all of the reported results (prior to publication). These results will be generated quarterly and will present performance measures and comparisons of individual member hospitals to an average "benchmark" as well as to other unidentified peer group hospitals.

Phased Approach to Making Measurements

In addition to the conceptual model of the measurement system and identification of the key system users, we believe that we must use an iterative phased approach that will allow us to begin making measurements while we continue learning "how best to make these measurements." This iterative approach will also allow us to move forward at varying rates in different organizations and even regions of the country. This is based on our firm belief that before one can expect to demonstrate improvements in any of the outcome measures associated with the CIS technology, we must first demonstrate that the key system users are actually using the information systems. Similarly, we believe that before we can expect to be able to measure any system use, we must be able to demonstrate that the requisite systems are in place and available to our key users. Therefore, we propose a three-phase iterative approach to beginning the measurements:

1. Phase I will consist of the measurements required to demonstrate "availability" of the systems.
2. Phase II will consist of the measurements required to demonstrate "use" of the systems.
3. Phase III will consist of the measurements required to demonstrate the effect of these systems on various performance measures that are often associated with IT use.

The first step in our research plan was to host a conference (November 11-12, 2004, Toronto, Canada) that would bring together people from a wide variety of stakeholder groups. In order to define the metrics, we needed to generate a consensus from many perspectives as to what was important to measure and how the measures should be calculated. On an ongoing basis, it is envisioned that these metrics would evolve and become much more comprehensive and complex; however, it is critical that the early-stage metrics be meaningful and feasibly generated from data that were clear, concise, and accessible.

The first day presentations included input from researchers, hospitals, integrated regions, consulting companies, vendors, and community care agencies. These presentations demonstrated first hand the strategy and the implementation of many information system initiatives throughout North America. The second day of the conference focused on measurement, highlighting the need to define a strategy and then implement a measurement and evaluation plan that reinforces that specific strategy. Then, post-implementation, the metrics can outline in

detail both the successes and the failures. Finally, the conference ended with placing the participants into breakout groups to work on defining the indicators that we will begin to track. The goal of our conference was to arrive at a consensus regarding the origin of these indicators. These results from our conference are presented below.

IT Costs

IT costs can be divided into four basic categories (adapted from van Bemmel and Musen [16]).

Hardware

This is all of the equipment necessary for data input, processing, communication, and archiving (eg, personal computers, servers, routers, network cabling or wireless access points, and storage devices). One should also factor in the equipment necessary to insure system reliability, including battery backup systems, off-site data storage and fail-over systems, and even on-site emergency power generators. This equipment could be purchased, rented, or leased. These costs should include the initial purchase price, the expected amortization period (usually 3-5 years), depreciation, and maintenance and operating costs. (Operating expenses generally include items such as computer storage tapes and disks, paper, printer cartridges, and so on. While often considered a small part of the total IT costs, a general rule of thumb is that the yearly cost of a printer, including depreciation, maintenance, and operating expenses, is roughly equivalent to the original purchase price.)

Software

This includes all of the software required to keep the organization functioning. This should include both system software such as the operating systems, database management systems, network operating systems, data communication software, and compilers (in the event that the organization is developing their own applications), along with the application software such as the results review, provider order entry, clinical documentation, admit/discharge/transfer, registration, scheduling, and billing. This software may be purchased, rented, or leased. These costs should include the initial purchase price or development costs, as well as ongoing maintenance contracts or costs (often one third of the original purchase or development costs).

Personnel

This represents all of the people (both central, assuming a local hospital is part of a larger organization, and local) required to keep the systems working, including management, developers, implementers, technicians, and those charged with system and application maintenance. In addition, one should factor in the costs of the people charged with providing initial training (both the cost of the trainers as well as the time spent by the clinicians away from their jobs) and ongoing support to the clinicians (eg, help desk operators). An initial estimate of this number could be the number of full-time equivalents (FTEs) in the information technology department along with an average overhead cost associated with FTEs in your organization. (Overhead costs include such items as financial and personnel management, furniture, and telephone and mail services and are usually calculated as a percentage, roughly 10-40% of each individual's salary and fringe benefit cost.)

Space

This number should reflect the costs to purchase, maintain, and manage the space or real estate required to house all of the personnel and equipment associated with the IT department. In addition to the purchase price, rent, or leasing fees and their associated amortization and depreciation costs, one should also factor in the costs of providing heat, light, and cleaning services within these areas. An initial estimate of this number could be the total number of square meters taken up by the IT department. Clearly, the cost of this space will depend greatly on whether it is located within the hospital or at an off-site facility. It will also vary depending on the use of the space, for example, space for personnel probably costs significantly more than the space required to store backup disks or tapes.

Any specific measures developed or selected should be capable of taking into consideration at least the following three main methods an organization might use to obtain its IT solutions: (1) buy it from vendors, (2) build it themselves, (3) or outsource the work. In each of these three modes, one would expect that some of the cost categories would increase while others would decrease. For example, if you buy a system from a vendor, your software purchase costs should be higher, but your personnel software development costs would consequently be lower. Likewise, if an organization outsources their work, then you would expect to see significantly lower costs in all four of the IT investment categories, but one would have to add back in the cost of outsourcing the contract.

IT Cost Measures

In order to begin this process of measurement, reporting, and analysis with as much consensus as possible, members of IMPROVE-IT convened to work on identifying the first generation metrics. One interesting debate focuses around whether the cost indicators should be just that, an indicator, or an all-inclusive cost calculation similar to a balance sheet item. In the end, the agreement was to focus on the former for two reasons. First, a simple straightforward indicator will be easier to calculate, which will entice more hospitals to submit their findings. Second, and perhaps more importantly, our emphasis is to identify a statistical relationship between IT spending and changes in health outcomes. As such, the actual amount invested is not as important as an *indicator* that can be considered as a predictor—not only of overall spending, but, hopefully, of changes in outcomes as well.

After much deliberation and consultation, the following measures were selected as the first generation of indicators along the cost axis:

- Amount of money spent on IT hardware over the last year: This straightforward indicator deals with current ongoing investment in hardware. It is hoped that this indicator will reflect the commitment to ongoing investment in new technology.
- Amount of money spent on software by the organization over the last year: This second investment indicator deals with the ongoing software costs.

- Total number of people on IT staff—FTEs: The third investment indicator will incorporate the human resources needed to operate and manage the new technology. Once again, this indicator will provide insight into the amount of support required.
- Amount of space: The final investment indicator simply measures the space required (ie, office space in square feet) to house the IT personnel and hardware for the organization.

IT Infusion

IT availability can be defined as the existence of, and access to, the requisite technology to collect, store, display, and transmit patient-identifiable, structured, clinical data in electronic formats. Therefore, we must be able to identify whether health care institutions and their providers have access to various health IT components.

One such metric would be in the area of percentage of patients in a region who have their health data available in an electronic format. As measurement techniques become more sophisticated, the measure could be improved to estimate the "completeness" of each patient's health record, although at the present time the definition of a "complete" EHR is still not precisely defined. On September 1, 2004, the American Health Information Management Association, Healthcare Information and Management Systems Society, and The National Alliance for Health Information Technology announced the formation of a Certification Commission for Healthcare Information Technology (CCHIT). They were charged with creating an efficient, impartial, and trusted mechanism for certifying ambulatory EHRs and other health care IT products. On January 7, 2007, the Commission announced that an additional set of 18 ambulatory EHRs had been certified, bringing the total to 55 [17].

Much of the current IT research literature, and practice, has focused on measuring and determining the optimal hardware and software configurations. What the industry truly needs, however, is analysis focused on the *use* of these computerized information systems and how they can provide organization-wide benefits.The adoption of new IT in the health care industry involves more than hardware and software issues. We need the ability to accurately measure the degree of "infusion" (or system capabilities), availability, and use of various CIS features so that we can begin comparing CIS implementations from different vendors at different organizations. While others [18] have developed very technically oriented measures, we believe [19] that one should go beyond technical attributes and focus on the behavior of clinicians to really answer the question, "What information technology is available and how is it used?"

This is not as straightforward a calculation as it might appear at first glance. Many subjective decisions are made independently by hospitals and other providers before any data are captured or analyses produced. These subjective decisions, which relate to what to capture, how to calculate it, and how to make the analysis relevant, all affect the final product. Due to the complexity of the concept of infusion, there are numerous options and metric calculations that can be selected. If two organizations make a different decision, which is almost a certainty, even if they happen to call the measure by the same name, the possibility is very low that they will compare identical factors. As a result, a cooperative venture is a necessary condition for meaningful comparisons. Once these measures have been agreed upon, then, and only then, can standards and baseline benchmarks be employed industry wide.

The general consensus is that *availability* and *use* of IT are two distinct concepts, and, therefore, we have identified three measures for each of these two separate concepts.

IT Availability Measures

To measure availability, the first indicator is the number of clinical applications that are available to 50% or more of the clinicians in an organization. As a proxy, "available" is interpreted as clinicians who "have a login that allows them to access that part of the system." Examples of the types of clinical applications we considered to be key components included:

- computer-based provider order entry (CPOE)
- computer-based order communication
- MD-level admitting, discharge, and daily progress notes
- RN-level nurse charting
- clinical laboratory results review
- picture archiving and communication systems (PACS)
- admit/discharge/transfer systems
- clinical data warehouse
- scheduling
- billing
- patient registration

Various types of clinical decision support (based on the Clinical Decision Support Implementers' Workbook [20]) are available:

- proactive order sets
- preventive health maintenance reminders
- drug ordering alerts: drug-drug interactions, drug-allergy interactions, duplicate therapy
- access to online reference materials
- condition- or order-specific data displays
- support for complex clinical guidelines, protocols, or pathways

The second availability measure is the percentage of time the CIS was "available for use" by clinicians. We termed this as the percentage of system uptime. It should be calculated as follows:

% system uptime = 100 × (total time – scheduled downtime – unscheduled downtime) / total time

Where:

- Total time is the total number of minutes in a day times the number of days over which the measure is taken.
- Scheduled downtime includes all scheduled reasons for system unavailability, including system upgrades, routine hardware maintenance, system backups, etc.
- Unscheduled downtime includes all unscheduled reasons for system unavailability, including power outages, equipment failures, software lockups, etc.

The third availability measure is total number of unique patients with some type of clinical data available in the clinical repository. If possible, it would be better to factor in the number of years of data that each of these patients has available, which would allow us to calculate an availability measure of the total number of patient-years of clinical data available to clinicians. Another relatively simple proxy for this measurement could be the amount of disk space taken up by the clinical data contained in all the clinical systems. In the end, a simple count of unique patients was selected.

IT Use Measures

IT use can be defined as actual hands-on employment of information systems by patients, providers, and those involved in population health. At the end-user level, this equates to actual use of various applications such as clinical results review or provider order entry. At the aggregate level, usage can be measured by the number of clinicians who routinely use the system to enter and review patient-level data.

The first CIS use measure is percentage of clinicians with an active user ID / password combination who actually log in to the system more than one time each day. We considered several other methods of "normalizing" the number of user log-ins, including number of log-ins per occupied bed and mean number of unique log-ins per individual patient. Once this measure reaches a uniformly high level, that is, the vast majority of institutions have well over 90% of their clinicians logging in each week, then we would consider revising this measure to reflect the mean percentage of all clinical applications available to each clinician that the clinician actually utilizes during the week or month.

For the second use measure, we selected percentage of patients with a completed chart (as defined by all needed data signed by all the appropriate clinicians) within 24 hours of their hospital discharge or outpatient visit.

For the final use measure, we focused on application-specific use measures. Here the goal would be to add one or more of these measures each year as our focus on the key clinical applications changes over time. Currently, there is tremendous emphasis on the use of computer-based provider order entry to reduce the number of errors in the ordering process; therefore, we chose the percentage of all orders entered directly by the person responsible for the patient's care (who could be an MD, physician's assistant, nurse practitioner, and the like) [21]. In subsequent years, we could easily imagine including, for example, the following:

- percentage of patients with a log-in to their personal health record who actually logged on in a given month
- percentage of clinicians who dictate their clinical notes (which currently requires the additional step of human transcription) rather than enter them directly via the keyboard
- overall percentage of clinical alerts or reminders that are overridden by clinicians

Health Performance

State-of-the-art IT can potentially help clinicians and ancillary personnel to improve the overall care delivery process, which should lead to improvements in health outcomes [22]. These improvements will not occur unless there is a concerted effort to improve the process itself. Evaluating the impact of advanced IT on the health care delivery system requires not only standard measures, but the measurements must also demonstrate that the IT led to, or helped lead to, the observed clinical outcome. In other words, one must be able to infer a potential relationship between the use of the IT and the observed measure. Identifying these relationships can be difficult. In addition, our objective is to use, as much as possible, available health outcome or process measures that are already being used in other clinical quality, safety, and effectiveness evaluation practices (such as various Balanced Scorecard initiatives [23]). The following example measures attempt to document various aspects of the IT evaluation framework outlined above.

We have chosen the measures developed by the Center for Medicare and Medicaid Services of the United States as part of their National Voluntary Hospital Reporting Initiative to represent this area. While we recognized that these "process" measures do not represent actual health outcomes, we felt that in an effort to begin this work, it was most important to select measures that virtually all US hospitals were already making. In the first release of this measure, hospitals were asked to report their performance in three areas of care:

1. acute myocardial infarction (AMI): In the United States, approximately 1 million people suffer an AMI each year, making it one of the leading causes of hospital admission for patients age 65 and older.
2. congestive heart failure (CHF): CHF is the most common hospital admission diagnosis in patients aged 65 or older, accounting for more than 700000 hospitalizations across the United States each year.
3. community acquired pneumonia (CAP): This causes 4 million episodes of illness and nearly 1 million hospital admissions each year.

Within each of these focus areas, the Center for Medicare and Medicaid Services identified two to five specific measurements that hospitals were asked to report [24]. Therefore, our complete list of health outcome measures is as follows.

AMI:

- percentage of patients hospitalized with AMI who receive their initial treatment 30 min after arrival at the hospital
- percentage of patients who receive their percutaneous coronary intervention within 120 min of hospital arrival

CHF:

- percentage of patients hospitalized with CHF who receive their left ventricular assessment within 30 min after admission to the hospital
- percentage of patients hospitalized with CHF who were prescribed an angiotensin converting enzyme (ACE) inhibitor prior to discharge

CAP:

- percentage of patients hospitalized with CAP who receive their initial dose of antibiotics within 4 hours after admission to the hospital
- percentage of hospitalized pneumonia patients 65 years or older who were given a pneumococcal vaccine, if indicated, prior to discharge
- percentage of hospitalized pneumonia patients who have an arterial blood gas drawn, or who are monitored using pulse oximetry, within 24 hours of hospital arrival

Currently over 90% of the hospitals in the United States are reporting these measures. We anticipate being able to use the values reported through this voluntary reporting initiative as our proxy indicator of health outcomes for each hospital.

Finally, we realize that it is often difficult to identify a specific link or relationship between a specific CIS feature and many of these high-level process measures, but we cling to the belief that better, more accessible information, like that provided by a state-of-the-art CIS, should lead to better and more rapid clinical decision making. These improvements in decision making, or perhaps a simple reminder that a decision needs to be made, should in turn lead to improvements in these process measures.

Measurement Going Forward

In order for the measurement to be meaningful, in addition to identifying the specific calculations, the size and type of hospital and the type of IT implemented must be consistent within peer groups. Therefore, each member hospital will need to be categorized on the following three factors:

1. number of beds
2. community care versus academic center
3. type(s) of IT

- physician/provider order entry (POE)
- electronic health record (EHR)
- clinical decision support (CDS)
- clinical data repository (CDR)
- ancillary systems interoperability (with internal systems such as labs, pharmacy, diagnostic imaging, emergency department triage systems)

This will result in much more meaningful comparisons as hospitals will be analyzed with respect to other similar, or peer group, hospitals.

Certainly, the specific outcome measures (either in terms of efficiency or effectiveness) that are calculated should relate in detail to the type of IT implemented (and being measured). The health outcome metrics presented above are intended to demonstrate the potential effect from any number of IT interventions. For example, the benefits for AMI patients could be the result of information delivery improvements related to the successful adoption of POE, EHR, CDS, CDR, or ancillary systems (such as emergency department IT). Other likely outcomes that must be further defined in order to relate to a designated IT system include the following:

- reduced length of stay
- lower readmission rates
- lower mortality
- reduced adverse events
- fewer complications with comorbidities
- faster turnaround cycle
- reduced human resource (doctors as well as other in-hospital staff) costs
- reduced diagnostic imaging costs
- reduced materials and supplies costs
- lower overall hospital costs (net of IT investment)

The following three steps emanate directly from the first generation metrics presented herein.

1. Establish national and international benchmarks for all common evaluation measures: Recruitment of membership is ongoing. We will use member data to develop a set of national and international cost, infusion, and effectiveness benchmarks. Benchmarks will be created to identify the 10th, 50th, and 90th percentiles of performance based on similar peer hospitals.
2. Explore statistical relationships between measures to illustrate potential cause and effect relationships: During the statistical analysis, we will identify the different factors that affect the timing and the amount of benefit that one should expect from IT investment, as this will allow for better prediction and easier management of expectations. Once a model of IT valuation is created, one of the primary benefits is the awareness of the interrelationships that exist among the many characteristics of the organization or its particular subindustry category.

3. Develop a complete and overall quality index that measures true impact of effective information systems in the inpatient setting: This will be accomplished over time as the data quality improves and the level of statistical analyses becomes more sophisticated.

Conclusion

This research will study whether increased IT capabilities, availability, and use lead to improved clinical quality, safety, and effectiveness in the inpatient clinical setting. To reiterate, the logic underlying this hypothesis is as follows:

- Investment in IT inherently provides newer and more powerful technology and technological solutions.
- This improvement in “solution power” should then generate, and hence make available, “better” (more timely, valid, relevant, precise) information.
- Increasing the availability of this better information within the health care setting makes it more likely that decision makers will access or use this information to make better decisions.
- Finally, these better decisions should lead to results of increased efficiency (time and monetary gains) and effectiveness (improvement in measurable health outcomes across a variety of dimensions).

Consequently, it is anticipated that the mere act of identifying metrics, doing the calculations, and making the comparisons

will have a positive impact on effective IT utilization in health care. There has been much established in the management literature pertaining to the act of measurement and the effect its mere presence can have on an outcome [25]. In particular, the Hawthorne Effect describes the fact that people perform better when they are being watched or measured, at least in the short term. The creation of indicators will highlight the importance of IT and will motivate the member hospitals to improve their results. Even if the measures identified herein are not optimal, they still serve a very important purpose of starting the debate as well as being the first steps in evaluating what is working and what is not.

In addition to the quantitative initiative presented here, we believe that similar qualitative studies should be conducted on the state of clinical and administrative information exchange standards and on the "values" of potential users of these systems. While these qualitative estimates of progress will not be as easy to interpret, they provide at least a glimpse of the progress that the industry is making in these critical arenas.

Examples of the types of topics these qualitative reviews might address include:

- qualitative assessment of the legal climate relating to public access to relevant data sources
- patient privacy protections
- legal restrictions on sending/receiving various data types
- electronic signatures
- prescription transmission to pharmacies
- legal restrictions on sending laboratory results to patients
- requirements to submit data to centralized databases
- availability of unique provider identification (UPI)

Likewise, in assessing the values of key system users, one might delve into:

- qualitative assessment of the perceived value of using IT for patient care
- incentives to adoption
- number of insurance companies reimbursing physicians for use of e-visits

Well-documented effects of health-related IT on health and health care represent vital metrics for the advancement of IT deployment. The value of the infrastructure ultimately must be evaluated, perhaps using the six quality attributes defined by the Institute of Medicine (safety, timeliness, efficiency, effectiveness, equitability, patient-centeredness) as measurement axes [26]. Although benefits and costs of IT have been measured in limited settings, measurements on the effects herein envisioned, on a national or international scale, have never been made. To accomplish this, of course, we must first establish the critical measurements of system availability and use.

In summary, while we firmly believe that the implementation and widespread adoption of IT throughout health care has had and will continue to have a significant positive effect, little documented evidence supports this belief. The IMPROVE-IT project is intended to demonstrate the tremendous positive influence that IT is having on health care. Improving efficiency requires knowledge of current inefficiencies, and improving effectiveness requires an understanding of the measurable outcomes of health care. No process can be managed or improved without first understanding the current status (ie, evaluating inherent performance measures).

Acknowledgments

Partial funding support was provided by the Canadian Institutes of Health Research (CIHR) Strategic Training Initiative in Health Research (STIHR).

Conflicts of Interest

The Improve-IT Institute (www.improve-it-institute.org) has been formed, and is owned, by the two authors. Their goal is to further develop the ideas outlined in this manuscript and to begin making and reporting the results of these measurements on their website.

Multimedia Appendix 1

IMPROVE-IT Presentation, presented at Mednet 2006 (ppt) [PPT file (MS Powerpoint), 87 KB - jmir_v9i2e9_app1.ppt]

Multimedia Appendix 2

Quicktime Video Presentation, Kevin Leonard, Dec 2006 [MOV file (Quicktime), 51.8 MB -]

References

1. Leonard KJ. Information systems for healthcare: Why we have not had more success - the top 15 reasons. Healthcare Mgmt Forum 2000;13(3):45-51.
2. Kuhn KA, Giuse DA. From hospital information systems to health information systems. Problems, challenges, perspectives. Methods Inf Med 2001;40(4):275-287. [Medline: 11552339]

3. Simon SR, Kaushal R, Cleary PD, Jenter CA, Volk LA, Poon EG, et al. Correlates of electronic health record adoption in office practices: a statewide survey. J Am Med Inform Assoc 2007 Jan;14(1):110-117. [Medline: 17068351] [doi: 10.1197/jamia.M2187]
4. Sittig DF, Shiffman RN, Leonard K, Friedman C, Rudolph B, Hripcsak G, et al. A draft framework for measuring progress towards the development of a National Health Information Infrastructure. BMC Med Inform Decis Mak 2005 Jun;5(1):14 [FREE Full text] [Medline: 15953388] [doi: 10.1186/1472-6947-5-14]
5. Leonard KJ. Investment in information technology: the dilemma of comparing costs to benefits. Hosp Q 1998;1(4):53, 55-53, 56. [Medline: 10345308]
6. Leonard KJ. Critical success factors relating to healthcare's adoption of new technology: a guide to increasing the likelihood of successful implementation. Healthcare Q 2004;7(2):72-81.
7. Rosenstein AH. Measuring the benefits of clinical decision support: return on investment. Health Care Mgmt Rev 1999;24(2):32-43.
8. Frisse MC. The business value of health care information technology. J Am Med Inform Assoc 1999;6(5):361-367 [FREE Full text] [Medline: 10495096]
9. Agrawal A. Return on investment analysis for a computer-based patient record in the outpatient clinic setting. J Assoc Acad Minor Phys 2002 Jul;13(3):61-65. [Medline: 12362561]
10. Kuperman GJ, Gibson RF. Computer physician order entry: benefits, costs, and issues. Ann Intern Med 2003 Jul 1;139(1):31-39 [FREE Full text] [Medline: 12834316]
11. Wang SJ, Middleton B, Prosser LA, Bardon CG, Spurr CD, Carchidi PJ, et al. A cost-benefit analysis of electronic medical records in primary care. Am J Med 2003 Apr 1;114(5):397-403. [Medline: 12714130] [doi: 10.1016/S0002-9343(03)00057-3]
12. ; National Committee for Quality Assurance. Physician Practice Connections: A Summary of Draft Standards and Elements (Version 2). Washington, DC: National Committee for Quality Assurance; 2005.
13. ; Advisory Board Company. Clinical IT Benchmarking Survey. Washington, DC: Advisory Board Company; 2006:183-220.
14. Beard N. Investment in information technology and hospital efficiency: a panel on data analysis. Proceedings of the 27th International Conference on Information Systems (ICIS); Milwaukee, Wisconsin 2006.
15. Schoen C, Davis K, How SKH, Schoenbaum SC. US health system performance: a national scorecard. Health Affairs 2006:w457-w457. [doi: 10.1377/hlthaff.25.w457]
16. Van Bemmel JH, Musen MA. Handbook of Medical Informatics. Houten, The Netherlands: Springer; 1997.
17. CCHIT Certified Ambulatory Electronic Health Record (EHR) Products by Product. cchit.org. URL: http://www.cchit.org/certified/2006/CCHIT+Certified+Products+by+Product.htm [accessed 2007 Feb 17] [WebCite Cache ID 5Mj0te92n]
18. Peel V. Evaluating the cost effectiveness, impact and value of hospital IT/IS implementations - lessons learned in the United Kingdom. In: Yeoh E, editor. Proceedings of the Hospital Authority Convention 1997 - Re-inventing Health Care for the 21st Century. Hong Kong: Hong Kong Health Authority; 1997:67-75.
19. Zmud RW, Apple LE. Measuring technology incorporation/infusion. J Prod Innov Manag 1992;9(2):148-155. [doi: 10.1016/0737-6782(92)90006-X]
20. Osheroff JA, Pifer EA, Teich JM, Sittig DF, Jenders RA. Improving Outcomes with Clinical Decision Support: An Implementer's Guide. Chicago, IL: Healthcare Information and Management Systems Society; 2005. URL: http://www.himss.org/cdsguide [accessed 2007 Feb 17] [WebCite Cache ID 5Mj13ChF4]
21. Sittig DF, Thomas SM, Campbell E. Consensus Recommendations for Basic Monitoring and Evaluation of In-Patient Computer-based Provider Order Entry Systems. Proceedings of Information Technology and Communications in Health (ITCH) February 15-18, Victoria, BC, Canada 2007.
22. Werner RM, Bradlow ET. Relationship between Medicare\'s hospital compare performance measures and mortality rates. JAMA 2006 Dec 13;296(22):2694-2702. [doi: 10.1001/jama.296.22.2694]
23. Castañeda-Méndez K, Mangan K, Lavery AM. The role and application of the balanced scorecard in healthcare quality management. J Healthc Qual 1998 Jan;20(1):10-13. [Medline: 10177013]
24. The Hospital Quality Initiative (HQI). URL: http://www.cms.hhs.gov/quality/hospital [accessed 2007 Feb 17] [WebCite Cache ID 5Mj1BUvoL]
25. Mayo E. The Social Problems of an Industrial Civilization. Boston, MA: Graduate School of Business Administration, Harvard University Press; 1945.
26. ; 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 Academy Press; 2001.

submitted 22.11.06; peer-reviewed by B Raymond, D Thompson; comments to author 04.01.07; revised version received 17.02.07; accepted 26.02.07; published 04.05.07.

Please cite as:
Leonard KJ, Sittig DF
Improving Information Technology Adoption and Implementation Through the Identification of Appropriate Benefits: Creating IMPROVE-IT
J Med Internet Res 2007;9(2):e9
URL: http://www.jmir.org/2007/2/e9/
doi:10.2196/jmir.9.2.e9
PMID:17513287

Viewpoint

Design and Evaluation in eHealth: Challenges and Implications for an Interdisciplinary Field

Claudia Pagliari, PhD

Division of Community Health Sciences (General Practice Section), University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Claudia Pagliari, PhD
University of Edinburgh
eHealth Interdisciplinary Research Group
Division of Community Health Sciences
20 West Richmond Street
Edinburgh EH89DZ
United Kingdom
Phone: +44 131 650 9464
Fax: +44 131 650 9119
Email: claudia.pagliari@ed.ac.uk

Abstract

Much has been written about insufficient user involvement in the design of eHealth applications, the lack of evidence demonstrating impact, and the difficulties these bring for adoption. Part of the problem lies in the differing languages, cultures, motives, and operational constraints of producers and evaluators of eHealth systems and services. This paper reflects on the benefits of and barriers to interdisciplinary collaboration in eHealth, focusing particularly on the relationship between software developers and health services researchers. It argues that the common pattern of silo or parallel working may be ameliorated by developing mutual awareness and respect for each others' methods, epistemologies, and contextual drivers and by recognizing and harnessing potential synergies. Similarities and differences between models and techniques used in both communities are highlighted in order to illustrate the potential for integrated approaches and the strengths of unique paradigms. By sharing information about our research approaches and seeking to actively collaborate in the process of design and evaluation, the aim of achieving technologies that are truly user-informed, fit for context, high quality, and of demonstrated value is more likely to be realized. This may involve embracing new ways of working jointly that are unfamiliar to the stakeholders involved and that challenge disciplinary conventions. It also has policy implications for agencies commissioning research and development in this area.

(J Med Internet Res 2007;9(2):e15) doi:10.2196/jmir.9.2.e15

KEYWORDS

Medical informatics; software development; evaluation; interdisciplinarity

Aims and Origin of This Article

This paper represents a personal viewpoint based on a nonsystematic review of the literature and the experience of observing and participating in the design, evaluation, and analysis of health informatics interventions. It originated as a briefing document for members of a multidisciplinary team of clinicians, researchers, and software designers, which was designed to foster shared understanding and plan a program of formative and evaluative work. The paper draws on existing literature advocating interdisciplinary methods in medical informatics but focuses on generating a dialogue between software developers and researchers working in this area.

The article begins by considering the increasing heterogeneity of the field, the need for multiple research perspectives, and the implications of scientific subcultures; it discusses the importance of research for ensuring that new eHealth technologies are adopted and effective; it highlights common concepts and methods in software design and health services research; and it then considers the benefits, challenges, and facilitators to interdisciplinary collaboration.

For the purposes of this paper, the term *eHealth* is used broadly as a synonym for health informatics or medical informatics and health services research for health technology assessment and health systems research.

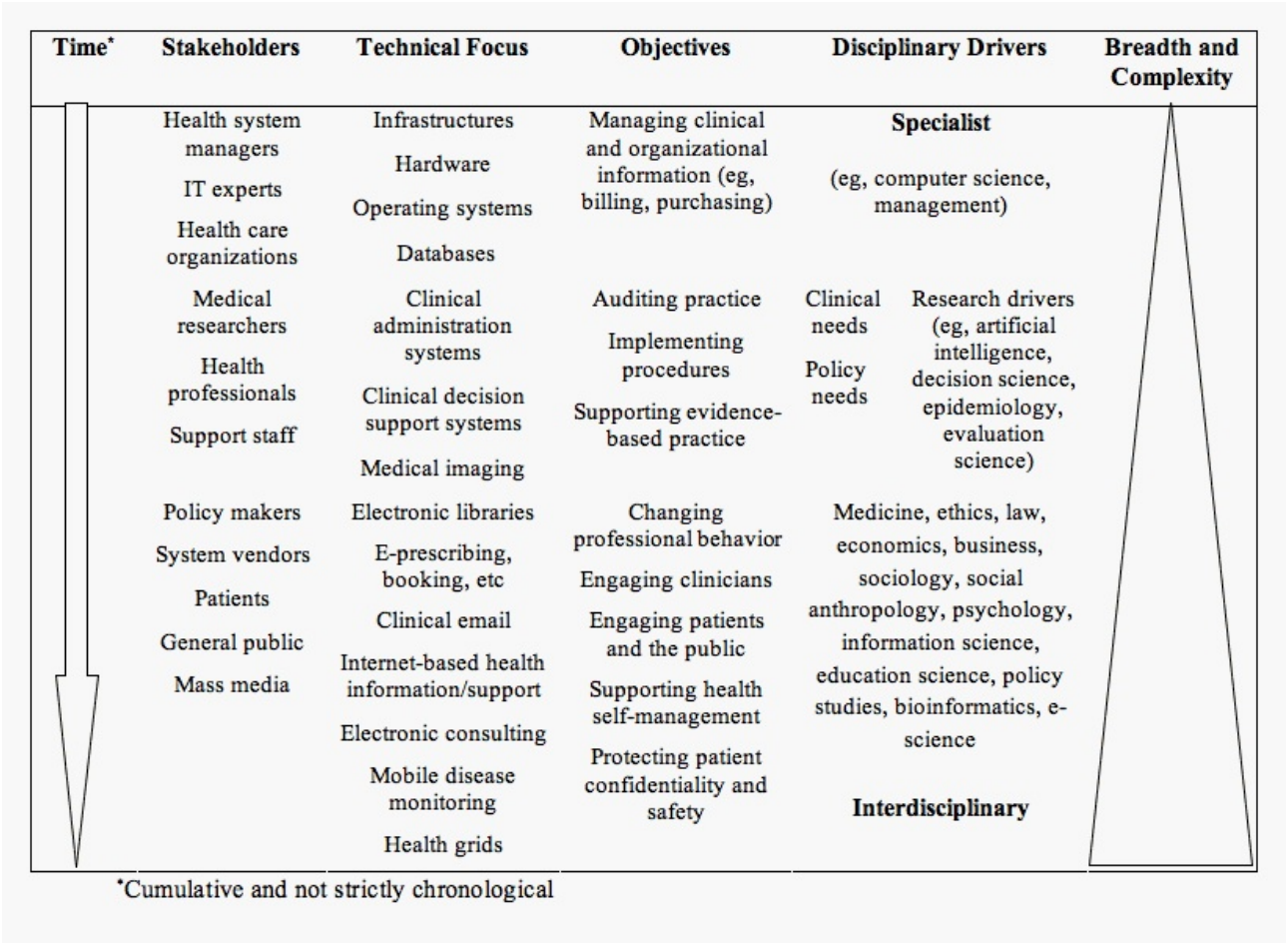
Out of the Basement: Changing Stakeholders in Medical Informatics

Not so long ago, medical informatics was largely the preserve of computing professionals and managers due to its focus on aspects of information technology "hidden" beneath the surface of health care organizations, such as operating systems, architectures, and databases. While epidemiologists were quick to harness the potential of electronic patient records for research and disease surveillance, it was the growth in practice-based computing during the 1990s that increased awareness of information technologies among clinical stakeholders and saw their gradual integration into the processes of care. Among the general public, awareness of eHealth, as the field is becoming known [1], has burgeoned since the turn of the 21st century, paralleling access to the Internet and the proliferation of Web-based health and lifestyle resources. This is reflected at the policy level, where governments have become increasingly interested in the potential of information and communications technologies to improve the organization and delivery of health services and to support patient empowerment for self-care [2]. In the United Kingdom, for example, the National Health Service (NHS) National Programme for Information Technology is gradually bringing the health service into people's homes via initiatives such as NHSDirectOnline and NHSHealthSpace, which offer not only information but also opportunities for electronic consulting and personal health care organization (eg, records, appointments) [3].

Reflecting this societal trend, academic involvement in medical informatics has become ever more interdisciplinary, with growing participation by the social, economic, and legal sciences (eg, around managing change, ethics, and cost-effectiveness) and the emergence of translational fields such as bioinformatics that promise to challenge existing medical models. At the same time, the boundaries between scientific, policy, and commercial areas of research and development are becoming grayer, as academia and industry respond to government funding opportunities and the policy community responds to emerging evidence and new technologies.

Growing use of the term *health informatics*, in preference to medical informatics, also reflects a shift toward inclusiveness. Figure 1 represents this shifting landscape in terms of the stakeholders, technical focus, disciplinary drivers, and objectives of medical informatics practice and research. It is not intended as a comprehensive chronological account of the field's evolution, although the domains reflect observations from previous analyses of the literature [1]. A key change has been the increasing breadth and complexity of the field not only in terms of new technologies but also the perspectives that are being brought to bear in planning, understanding, and evaluating these technologies.

Figure 1. The increasing breadth and complexity of eHealth in terms of the stakeholders, technologies, objectives and disciplines involved



Scientific Subcultures and Their Implications for Interdisciplinary Working

The increasing heterogeneity of the eHealth field raises challenges for interdisciplinary working and the translation of research to policy and practice. These challenges have to do with the management of nonshared concepts and languages and the values ascribed to different forms of scientific and technological endeavor, within what may be termed the knowledge economy of eHealth.

Despite popular stereotypes about "the" scientific paradigm, different disciplines have evolved uniquely over time and have their own theoretical or applied stance, criteria for appraising quality, and their own ways of working. Although we agree on basic principles of objectivity and methodological adherence, the way in which we see the world and approach new research problems is affected by a host of contextual and historical factors that are specific to individual disciplines, making truly interdisciplinary working difficult to achieve [4].

Unpicking all these factors is beyond the scope of a single paper; however, a useful guide is to be found in the comparison of two important areas of activity that are central to modern eHealth—namely, software design and health services research (HSR). The reason for concentrating on these is that managing their relationship is fundamental to ensuring that eHealth innovations achieve their potential to improve the quality, efficiency, and safety of patient care. This paper focuses on the software development process, with particular reference to user-centered design methods. Its origins lie in the author's growing appreciation of the nature, value, and limitations of evaluation methods used in the software engineering community and the lack of awareness of these among health services researchers evaluating eHealth resources. While several high-profile documents have explored potential synergies between HSR and the broad field of medical informatics, and these have undoubtedly contributed to quality improvements in some areas, their potential impact across the wider eHealth landscape is far from being realized [5-12]. In practice, many eHealth software developments, and the HSR projects associated with them, take place in the context of short contractual episodes, where neither developers nor health services researchers have the time or incentive to engage in cross-disciplinary learning. As a result, developers and researchers of eHealth regularly work in parallel universes, each regarding the other's domain of activity as separate and neglecting the potential for useful interaction.

The Need for More Research in Development

Although developing technical solutions remains central to medical informatics, recent years have seen a growing emphasis on identifying and resolving barriers to implementation. Particular attention has been devoted to understanding so-called people and organizational factors, such as stakeholder resistance to change and the appropriate integration of new technologies into work patterns [13,14]. Two key themes have emerged from this discourse, which have direct relevance for the potential effectiveness of eHealth innovations:

1. The clinical appropriateness and usability of eHealth technologies have been compromised by insufficient end-user engagement in the design process.
2. The effectiveness of emerging eHealth technologies in improving the processes or outcomes of health care is unproven.

To consider the first theme, while there is general consensus among software designers on the importance of engaging users in software design and testing, commercial drivers and a historical focus on product development have meant that this has often been inadequate in the past, resulting in top-down developments whose problems may only emerge after rollout. The health care sector has been particularly prone to such problems in recent years, and there are numerous examples of potentially useful systems that have failed or been abandoned due to unanticipated technical, human, or organizational issues [15-17]. Design flaws can affect the ease of use and reliability of systems and may even be dangerous, creating ill-feeling and reducing clinicians' willingness to use emerging systems, software, and hardware in practice [18,19]. Even seemingly minor problems with usability or conceptual fit can destabilize the implementation of otherwise highly engineered and valid technologies. The discussion that follows illustrates how developers are rising to this challenge.

To consider the second theme of eHealth technologies being unproven, while research in this area is burgeoning, it remains a fact that there is little reliable evidence to demonstrate the measurable impact, risks, or cost-effectiveness of eHealth innovations, except in a modest number of application areas [20-22]. This creates uncertainty and hence a reluctance on the part of clinicians and policy makers to implement such technologies. Where rigorous research designs have been employed, this has often been in the context of academic studies in which the future sustainability or generalizability of the products being evaluated cannot be assured. Indeed, a recent systematic review of health information technologies demonstrated that of 257 published evaluations, a staggering one quarter emanated from four academic institutions that implemented internally developed systems, while only nine reported on commercially developed systems [22].

Tackling these problems requires the application of joint thinking between practitioners in the two fields so as to ensure high-quality, user-informed products of demonstrated effectiveness. However, cultural divides between the traditional software developer and health researcher communities have inhibited this process.

An Evolutionary Snapshot

Software development represents an application of computer science, a field rooted in engineering and mathematics. Although it has drawn on philosophy (eg, semantics, logic) and social science (eg, human-computer interface research, social technology studies), its historical focus has been on building machines and the software they require, albeit with ever more

complex digital innovations such as the Internet and intelligent agents. This focus on product development has led to a close alliance with the business and service sectors, and, although basic science is highly valued, there has been an understandable emphasis on applied research and development within university curricula. Within the workforce itself, economic drivers prioritize the production of resources that meet key functionality criteria and client-defined requirements within commercially viable time frames. Evaluation often takes a lower priority, and rapid application development using small convenience samples of users is common [23].

HSR is an interdisciplinary field concerned with the scientific study of the structure, processes, and effects of health services, technologies, and policies. This harnesses traditionally medical research approaches from epidemiology and clinical science, alongside the social and economic sciences, utilizing a mixture of quantitative and qualitative methodologies appropriate for the specific problem under investigation. It is closely allied to the evidence-based medicine movement, which holds that clinical practice should be driven by evidence of what interventions work best and for whom. As well as measuring impacts, such research is also about enquiry. For example, it may explore the needs of particular stakeholder groups or demographic patterns of health and health care utilization in order to identify the place of a potential new intervention or examine the reasons why an intervention is more easily adopted or more effective in different contexts. A defining characteristic of this field is the strong emphasis on methodological rigor. From randomized controlled trials to qualitative case studies, the focus is on detailed planning and recording of procedures and on transparent, theoretically informed participant sampling and data analysis. This area is less influenced by commercial drivers, although there is a strong emphasis on research that addresses health service policy needs.

Thus, software design is mainly concerned with developing interventions, and HSR, with evaluating them. But look closer and the reality is not so clear cut. In fact, much of HSR is geared toward informing the design of new interventions, including eHealth technologies, while rigorous software design encompasses evaluation processes that would be very familiar to health services researchers.

However, within these two communities there has been a mutual lack of awareness of each others' theoretical stance, motives, and modus operandi, exacerbated by differences in language, epistemologies, and the representation of concepts. This reflects the origins of the two disciplines and the funding environment, which place different expectations on eHealth design and research projects.

Compatibilities in Models and Methods of Software Development and Health Services Research

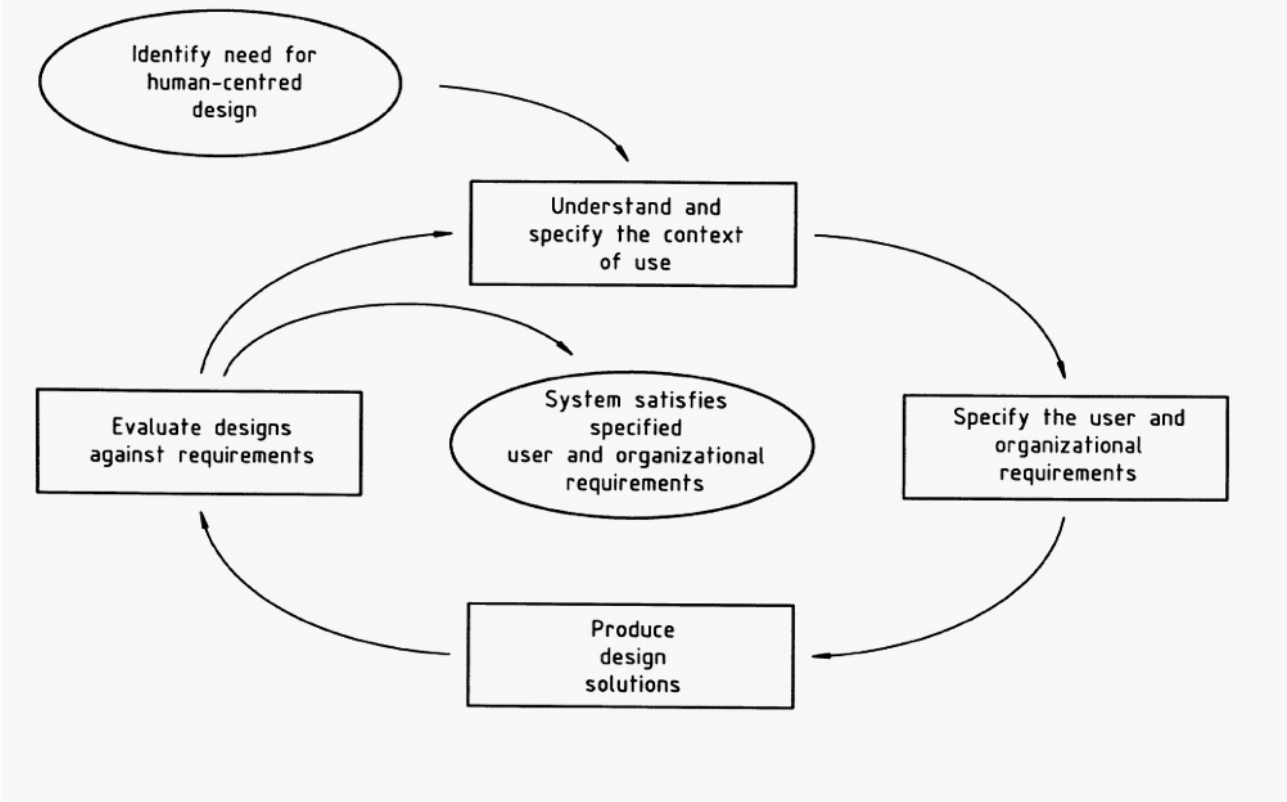
While various academic approaches have been applied to the study of software design and diffusion (eg, in the management literature), in the context of this paper the compatibilities between process models of software development and HSR are particularly relevant. These compatibilities illustrate the importance of exploration and evaluation for informing developments and quality improvements in both domains, the value of user engagement in this process, and the natural progression to assessment of the effects.

Lifecycle Models as an Exemplar from Software Engineering

Within software engineering, numerous models have been proposed to describe the process by which products should be designed and tested to ensure they are fit for purpose in the intended setting [24]. Particular parallels with HSR can be found in a category known as lifecycle models, the most common of which are the Waterfall, Spiral, and Star models, referring to the sequence and pattern of substages involved (Figure 2) [25-27]. Of these, the Spiral and Star models are frequently advocated due to their ability to cope with iteration and complexity, although in practice the more sequential Waterfall method is often used [28]. All of these illustrate the codependence of development and evaluation, while the Spiral and Star models emphasize iterative design. Although they have been slow to evolve, software design and development methodologies now almost universally include user-developer interaction for requirements determination, testing, and acceptance activities; indeed, this is a central feature of the Star Model. Figure 3 illustrates the model of user-centered design of the International Organization for Standardization (ISO) that is increasingly adopted when developing interactive systems [29]. The critical feature of this, and other approaches to user-centered design (or usability engineering), is the emphasis on determining users' needs of the system, understanding the context in which the system will be delivered, and designing products from the ground-up rather than based on developers' preconceptions or rigid procurement briefs. Such methods are being increasingly advocated, and their successful use is being reported in the medical research literature [30,31]. In some development settings, the user has taken a further step toward the center of the design process; for example, a paradigm employed in the defense sector uses software to directly involve users in developing their own problem-solving intelligent agents [32].

Figure 2. Key software lifecycle models: Waterfall [25], Spiral [26], Star [27] model

Figure 3. ISO 13407 standard for human-centered design processes for interactive systems



Phased and Iterative Models of Health Services Research

There is little awareness of software lifecycle models among typical health services researchers, yet these are highly compatible with phased approaches to drug development and the evaluation of complex interventions in health care, which emphasize the need for exploratory, explanatory, and pragmatic phases, as illustrated in Figures 4a and 4b [33]. Particular parallels can be seen in the stages of concept formation, needs assessment, and evaluation in the intended setting.

Figure 4a. Sequential stages in evaluation of complex interventions (after [33]). Similar steps are used in the evaluation of new drugs, from initial preclinical research through to postmarketing surveillance.

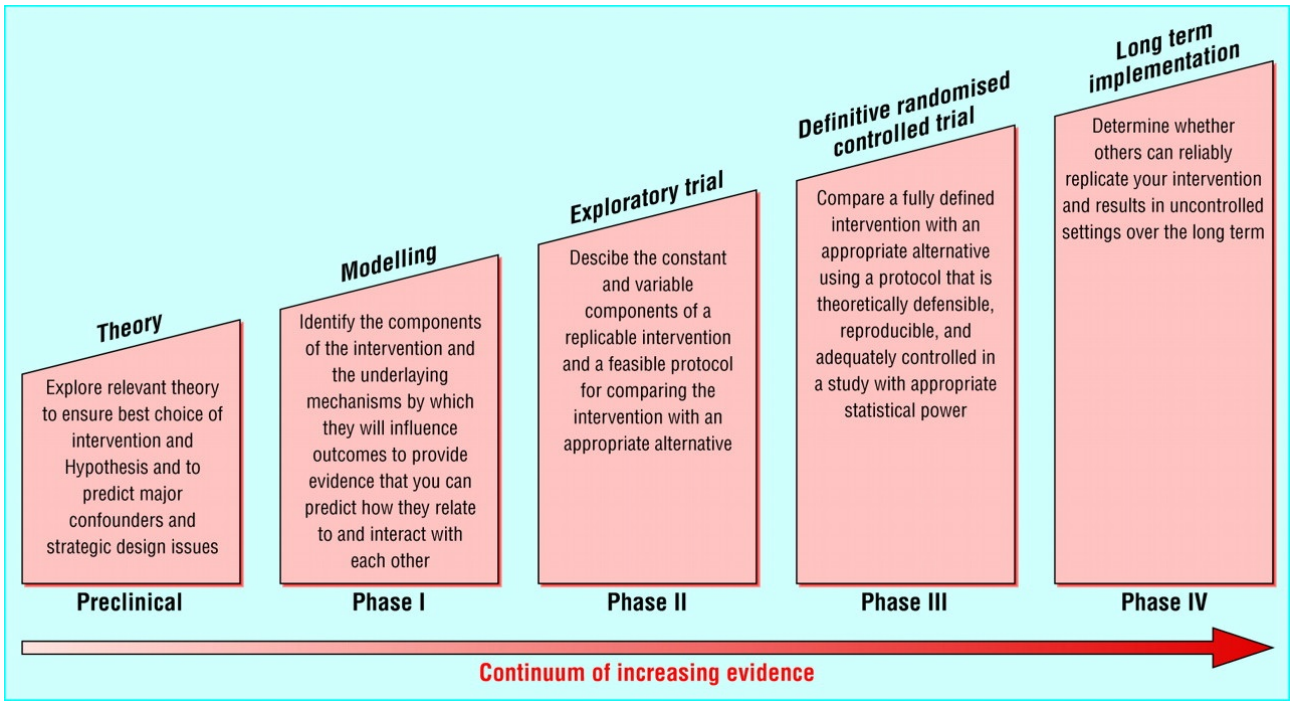
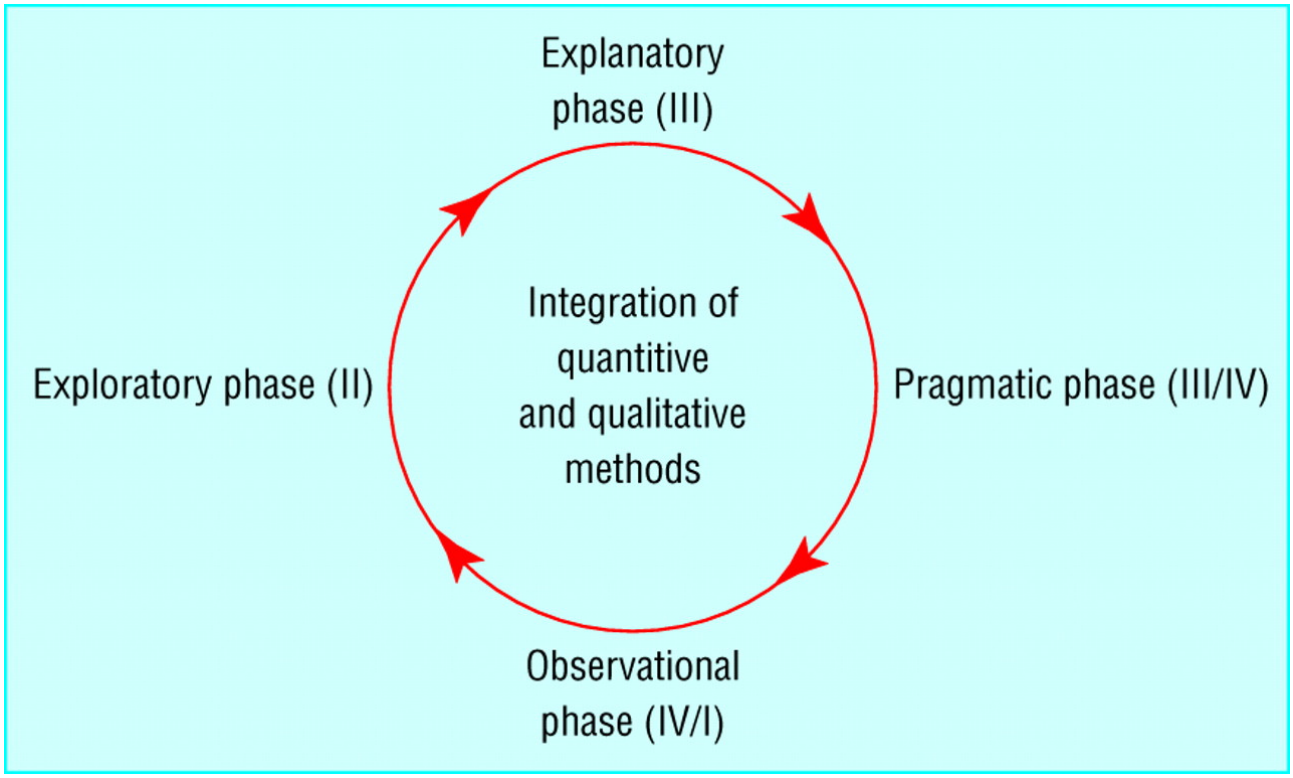


Figure 4b. Iterative view of complex intervention evaluation (after [33]). This recognizes that results from individual phases may prompt revisions and repetition.



Similarly, models of user-centered design bear a close resemblance to iterative HSR models such as Action Research [34] and Continuous Quality Improvement [35], examples of which are given in Figures 5 and 6. These also conceive of a cycle or series of cycles through which users' needs are assessed, interventions developed, problems identified, and changes made to the intervention or the management of its delivery. Indeed, these models are advocated within both the health care and software development arenas [36,37].

Figure 5. The Action Research Spiral (after [34])

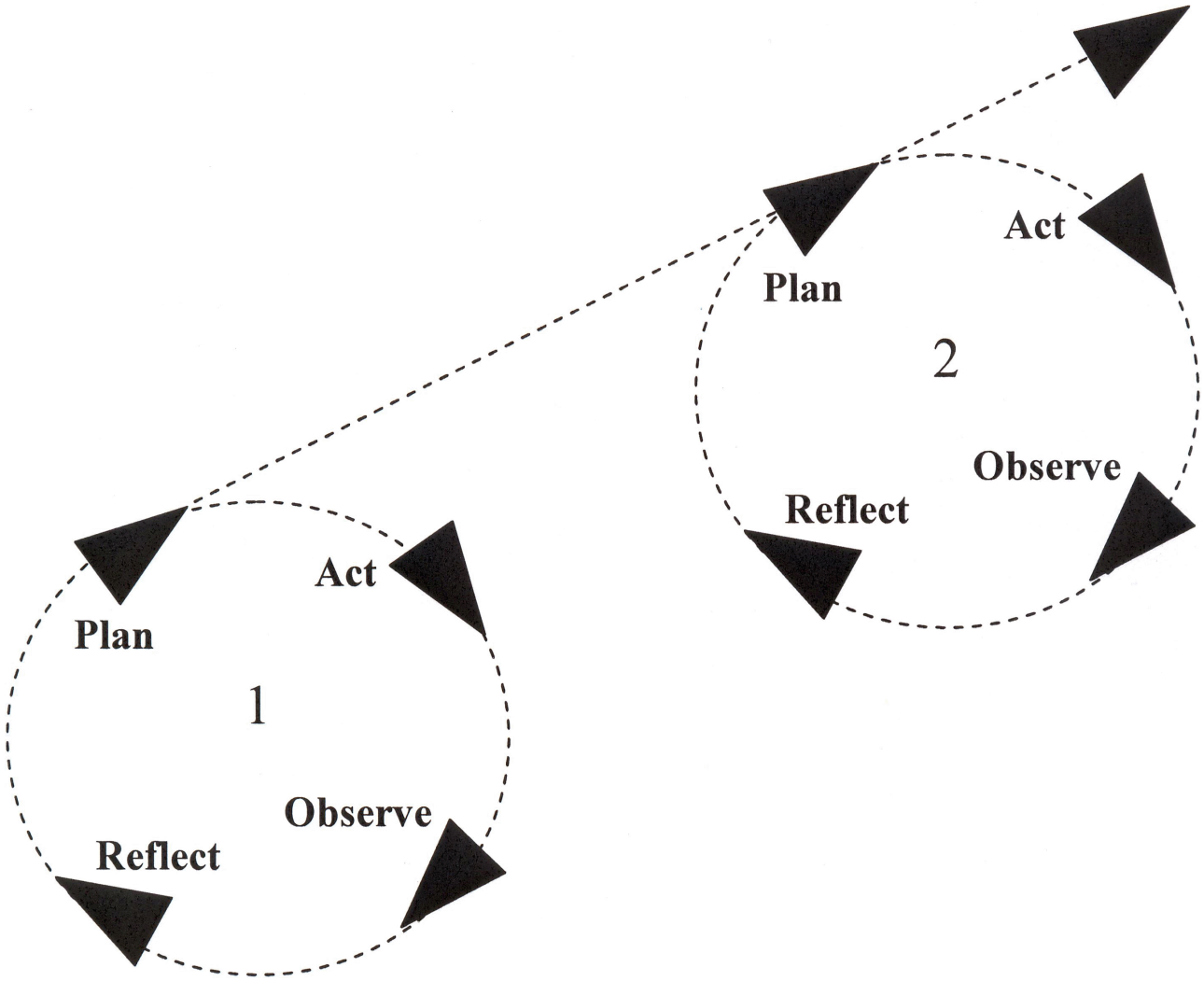
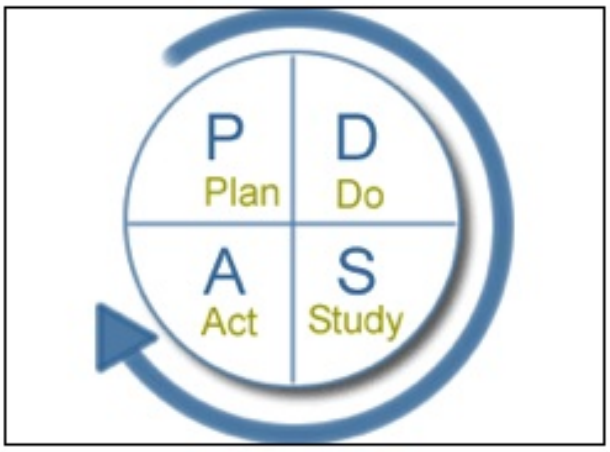


Figure 6. The "Plan-Do-Study-Act" process improvement cycle of total quality management (after [35])



Overlapping Research Techniques in Health Services Research and Software Design

As well as the parallels between overarching process models, there is considerable overlap in the precise research techniques used in both software development and HSR. While the terminology varies, user-centered design methods such as requirements gathering, observation of task walk-throughs, think aloud protocols, and group-based feedback are similar to HSR methods such as needs assessment, participant observation, semistructured interviews, and focus groups; indeed, these terms are regularly used to describe activities in information technology labs, although the way in which they are applied may be somewhat different. Examples of techniques used in

both fields are provided in Table 1 to illustrate areas of overlap and divergence. Differences in the portfolios of methods reflect the somewhat dissimilar (although overlapping) goals of evaluation within software engineering and HSR: the former focusing on optimizing product design and fitness for purpose, and the latter on exploring new phenomena, generating hypotheses, demonstrating impact, or informing policy. Most noteworthy is the absence of rigorous impact assessment (controlled trials) within the scope of software engineering, in contrast to its high status within HSR. An important differentiating feature not reflected in Table 1 is the heavy emphasis on theoretical sampling and meticulous time-intensive approaches to qualitative data analysis within HSR. This contrasts with the more rapid identification of needs and responses common in development projects and the often unstructured and iterative nature of the design process. Nevertheless, software engineering can also involve quantitative usability techniques that draw on cognitive psychology. When these are employed, it is often in a highly systematic manner, involving multiple measurements and theoretically based analysis, although the objectives are best met with depth studies of small numbers of users. These methods have great value for the understanding of human errors and information processing, and, although there is little knowledge of them within the HSR community, they are increasingly being reported in medically indexed journals [38-40].

Table 1. Examples of methods in user-centered design and health services research

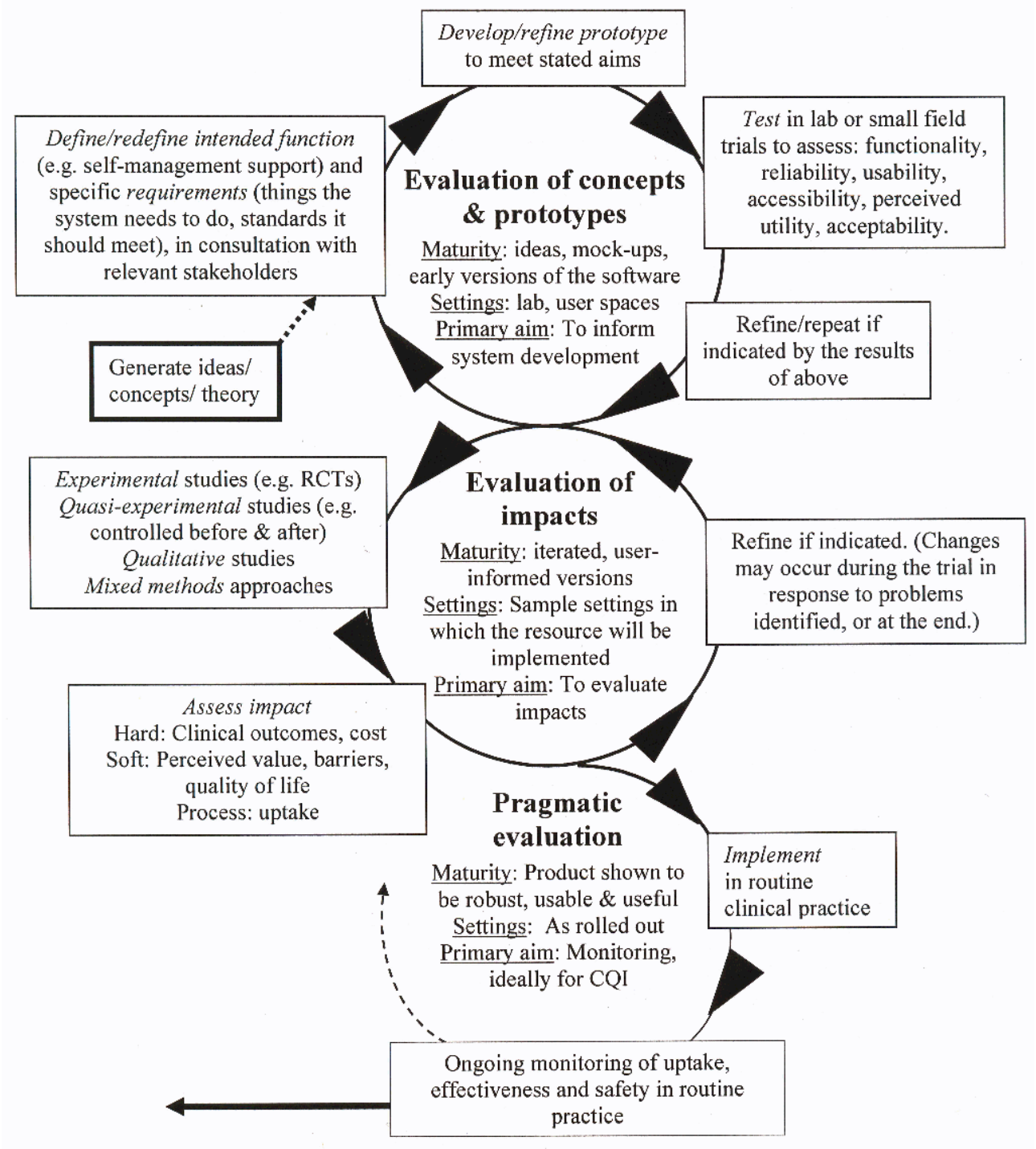
Software and Usability Engineering	Health Services Research
Needs Assessment (conceptual, formative)	**Needs Assessment** (conceptual, formative)
Requirements gathering: assessment of prototypes/simulations; user interviews	Interviews; document analysis; telephone or postal surveys; focus groups; observation; discrete choice simulations
Assessment (primarily formative)	**Assessment** (formative or summative)
Heuristic evaluation; cognitive walk-throughs; formal usability inspection; pluralistic walk-throughs; feature inspection; consistency inspection; standards inspection; guideline checklists; thinking aloud protocol; prototyping; co-discovery methods; question asking protocol; performance measurement; gaze tracking; ethnographic study / field observations; surveys; questionnaires; journaled sessions; self-reporting logs; remote usage observation; screen snapshots; blind voting; card sorting; archetypal research, action research	Experimental and quasi-experimental designs (eg, randomized controlled trial; controlled before and after study; interrupted time series; case control study; cost-benefit analysis) Qualitative outcomes assessment: rigorous qualitative data analysis using sociological methods (eg, ethnographic studies) Observation/exploration: remote (eg, epidemiological, records-based); direct (eg, participant or nonparticipant) Participative evaluation (eg, action research / continuous quality improvement)

Note: This is a non-exhaustive list drawing on several taxonomies that is designed merely to illustrate some of the common and distinctive techniques used in both disciplines.

Integrated Models in Medical Informatics

Within the interdisciplinary field of medical informatics, hybrid models have appeared that draw on both traditions, an example of which is offered in Figure 7. Importantly, there is a growing acceptance that evaluation should ideally be approached as a longitudinal process occurring through a series of overlapping and iterative stages relevant to the maturity of the technology in its lifecycle, from initial conception to rollout. Various authors have attempted to represent these stages and to provide taxonomies of research methods appropriate to each [41]; however, three broad phases of activity may be discerned. The first of these involves drafting new interventions based on an assessment of stakeholder needs and theory, and testing these with content experts and users to ensure they fulfil these needs and are technically robust (concept and prototype evaluation). The second involves assessing the impact of the innovations on the processes and outcomes of care in selected target settings, including hard measures such as efficiency, clinical status, cost, and error rates, softer measures of attitudes and satisfaction, and qualitative outcomes (outcomes evaluation). A third phase involves evaluating systems after rollout (pragmatic evaluation), for example, to assess variations in uptake, reported errors, technical problems, stakeholder satisfaction, or longer term impacts on process or outcome indicators. At each of these stages, the model should allow for the results of the research to inform continuous quality improvements. In practice, key stages are often neglected, reducing both the quality and adoption of new eHealth products.

Figure 7. An idealized framework for evaluating emergent eHealth resources at different stages of development and implementation



Benefits, Challenges, and Facilitators of Interdisciplinary Collaboration

Potential Benefits

The types of conceptual and methodological commonalities outlined above demonstrate that collaborative working between eHealth services researchers and software developers is both possible and appropriate. In addition to producing better and safer interventions [18,19,42], effective collaboration will strengthen the quality of evaluations and enhance the evidence base in this area, thus facilitating policy and purchasing decisions. While involvement in formal research may seem like a hindrance to developers, particularly in the commercial environment, economic as well as intellectual benefits may accrue by demonstrating that systems are effective, cost-effective, and safe, as well as technically robust, accessible, and usable [8]. Indeed, the value of so-called evidence-based business is being increasingly recognized in the technology sector [43]. At the same time, rigorous qualitative studies can demonstrate the acceptability and utility of new tools to users as well as features of the setting or implementation methods that may influence their adoption. For health services researchers, the ability to enter the world of the developer presents valuable opportunities to influence the scoping, design, and evaluation processes used to develop electronic health care

interventions that may then be subjected to clinical trials, thus ensuring conceptual fit and minimizing the risk of confounding by suboptimal functionality or usability [44]. It also offers a new skill set that may help researchers to recognize cognitive barriers to technology adoption and thus aid interpretation of descriptive or evaluative data. An added benefit for both groups is the increased ability to publish that comes from adopting systematic and replicable sampling and analytic methods in the course of user-centered design, thus facilitating dissemination to both audiences [45].

Challenges

There is a need to move the current agenda away from a state of parallel working, which is common in multidisciplinary projects involving the computing and medical sciences, to one of truly interdisciplinary working. This requires an appreciation of each others' terminologies, goals, and methods and the sharing of experiential learning about the benefits and limitations of alternative approaches. It also calls for the generation of a new breed of transdisciplinary experts familiar with the implementation of both skill sets and able to combine them in novel ways in order to achieve maximum value for eHealth research and development. Overcoming cultural and methodological divisions between disciplines represents a major challenge. There is a natural inclination to remain pure to concepts and methods that may have taken many generations to evolve, and questions arise around how far to take joint approaches before compromising each discipline's ability to demonstrate their specific expertise [4]. There is also a fundamental tension between the need to innovate, which may require conceptual leaps of faith and rapid developments, and the pressure to adopt methodologically robust standards of scoping, sampling, and evaluation that may be time-consuming and of questionable value at the early stages of prototyping. This can create antagonism and defensiveness in both camps, thus inhibiting potential synergies. Successful interdisciplinarity therefore requires the establishment of trust and mutual respect in addition to methodological pluralism, and this represents a challenge, particularly where the opportunity to become embedded in the other's world is not available. Importantly, different approaches will be appropriate for addressing different objectives in different settings and at different stages of software maturity, and a challenge for project leaders and commissioners is to develop a deep enough understanding of multiple methods to be able to tailor these appropriately. For example, controlled trials may be ideal for studying the impact of eHealth systems on measures of clinical outcome or efficiency, but they are poorly suited to exploring social, contextual, or technical barriers to adoption and certainly will have little to offer developers designing a new Web interface. Conversely, think aloud methods may be extremely useful for assessing the usability of a decision support tool but say very little about its clinical validity or effectiveness [46-48]. The value attributable to different forms of evidence thus varies depending on the context in which it is used, although adherence to high standards of data collection, analysis, and reporting is a universal objective. It should also be recognized that academic incentives favoring controlled studies (eg, research funding, impact ratings) may create a conflict for health services researchers wishing to engage in applied and collaborative projects.

Facilitators

The move toward more holistic training in medical informatics advocated by bodies such as the American Medical Informatics Association represents one step to achieving these goals [5,9], and there is evidence of a trend toward increasing pluralism in the objects of evaluation projects, which may signal a move toward greater interdisciplinarity [48]. Pockets of transdisciplinary working are emerging as eHealth becomes a target of research, for example, within academic units of human-computer interaction and science and technology studies, while the field of information science has a long tradition of research exploring socioeconomic and organizational influences on technology development and adoption, from which eHealth researchers and developers have much to learn [49]. Nevertheless, few individuals working on eHealth projects have received formal cross-disciplinary training and many are doing so as part of a broader portfolio of projects (often on short-term contracts), restricting their motivation to invest in learning the methods and modus operandi of their disciplinary counterparts.

There is a need to influence potential funders, who have traditionally held different expectations for design and evaluation projects in terms of expected outputs (eg, new products vs new knowledge) and methodologies (eg, user-centered design vs studies of clinical impact) and who may underestimate the value of unfamiliar approaches. Importantly, it is essential for those commissioning new eHealth products to be aware of the value of high-quality evaluation during the development process and to allow the time and resources for this to be built into the project. While research agencies are coming to recognize the need to pay attention to usability engineering and other software design methodologies when developing eHealth tools for research, the message of added value needs to be more widely communicated. This is particularly so in view of the revenue currently being invested in health-related websites, many of which are often of poor quality and unknown effectiveness [50], and the vast expenditure being devoted to eHealth technologies by governments worldwide [51]. Without this understanding, those commissioning products in this area will continue to be unwittingly complicit in the process of suboptimal design, while those commissioning evaluation will risk poor value for money if the questions asked are inappropriate or the research methods not suited to answering them [52-54].

Conclusions

Designing effective eHealth systems and services requires the application of expertise from diverse fields and will benefit from interdisciplinary collaboration. This may be eased by increasing familiarity with each others' terminologies, theoretical bases, and research methods, with the ultimate objective of achieving transdisciplinary working. There is sufficient overlap in the techniques and concepts employed within the software design and HSR communities to make this a reality, but realizing this requires the development of mutual trust and respect for each others' aims, epistemologies, and

contextual drivers, as well as a willingness to step outside traditional working boundaries. New funding strategies that recognize the value of alternative methodologies and of joint working between developers and evaluators are also called for.

This paper has merely scratched the surface of a wider debate on the value of interdisciplinarity for improving the quality and effectiveness of eHealth, although it is hoped that by highlighting the potential synergies between HSR and software development it will help to provoke constructive dialogue between these two communities. Maximizing the potential of eHealth also requires the involvement of a wider constituency of disciplinary experts, including social, management, and legal scientists, all of whom have a stake in the field. Interdisciplinary networks, such as the one managed by the author, offer one means of addressing this need.

Acknowledgments

The author would like to thank Professors Denis Protti, Jeremy Wyatt, and Peter Gregor for their constructive comments on the draft.

Conflicts of Interest

None declared.

Multimedia Appendix

PowerPoint slides of MedNet 2006 presentation [PPT file (MS Powerpoint), 1008 KB - jmir_v9i2e15_app1.ppt]

References

1. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005;7(1):e9 [FREE Full text] [Medline: 15829481] [doi: 10.2196/jmir.7.1.e9]
2. McConnell H. International efforts in implementing national health information infrastructure and electronic health records. World Hosp Health Serv 2004;40(1):33-37, 39-40, 50-2. [Medline: 15114922]
3. ; Department of Health. Delivering 21st Century IT Support for the NHS. National Strategic Programme. London, UK: Department of Health; 2002. URL: http://www.medgraphics.cam.ac.uk/Downloads/NHS_0007.PDF [accessed 2007 May 11] [WebCite Cache ID 5KF3I4EYY]
4. Weingart P. Interdisciplinarity: the paradoxical discourse. In: Weingart P, Stehr N, editors. Practising Interdisciplinarity. Toronto, ON: University of Toronto Press; 2000:25-42.
5. Corn M, Rudzinski KA, Cahn MA. Bridging the gap in medical informatics and health services research: workshop results and next steps. J Am Med Inform Assoc 2002 Mar;9(2):140-143 [FREE Full text] [Medline: 11861629] [doi: 10.1197/jamia.M0971]
6. Ammenwerth E, Brender J, Nykänen P, Prokosch HU, Rigby M, Talmon J; HIS-EVAL Workshop Participants. Visions and strategies to improve evaluation of health information systems. Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. Int J Med Inform 2004 Jun 30;73(6):479-491. [Medline: 15171977] [doi: 10.1016/j.ijmedinf.2004.04.004]
7. Gagnon MP, Scott RE. Striving for evidence in e-health evaluation: lessons from health technology assessment. J Telemed Telecare 2005;11(Suppl 2):S34-S36. [Medline: 16375790] [doi: 10.1258/135763305775124722]
8. Henderson J, Noell J, Reeves T, Robinson T, Strecher V. Developers and evaluation of interactive health communication applications. The Science Panel on Interactive Communications and Health. Am J Prev Med 1999 Jan;16(1):30-34. [Medline: 9894552] [doi: 10.1016/S0749-3797(98)00106-8]
9. Shortliffe EH, Garber AM. Training synergies between medical informatics and health services research: successes and challenges. J Am Med Inform Assoc 2002 Mar;9(2):133-139 [FREE Full text] [Medline: 11861628] [doi: 10.1197/jamia.M0974]
10. Mandl KD, Lee TH. Integrating medical informatics and health services research: the need for dual training at the clinical health systems and policy levels. J Am Med Inform Assoc 2002 Mar;9(2):127-132 [FREE Full text] [Medline: 11861627] [doi: 10.1197/jamia.M0973]
11. Grigsby J, Brega AG, Devore PA. The evaluation of telemedicine and health services research. Telemed J E Health 2005 Jun;11(3):317-328. [Medline: 16035929] [doi: 10.1089/tmj.2005.11.317]
12. Kaplan B. Evaluating informatics applications--some alternative approaches: theory, social interactionism, and call for methodological pluralism. Int J Med Inform 2001 Nov;64(1):39-56. [Medline: 11673101] [doi: 10.1016/S1386-5056(01)00184-8]
13. Kaplan B, Brennan PF, Dowling AF, Friedman CP, Peel V. Toward an informatics research agenda: key people and organizational issues. J Am Med Inform Assoc 2001;8(3):235-241 [FREE Full text] [Medline: 11320068]
14. Karsh BT. Beyond usability: designing effective technology implementation systems to promote patient safety. Qual Saf Health Care 2004 Oct;13(5):388-394 [FREE Full text] [Medline: 15465944] [doi: 10.1136/qhc.13.5.388]

15. Southon G, Sauer C, Dampney K. Lessons from a failed information systems initiative: issues for complex organisations. Int J Med Inform 1999 Jul;55(1):33-46. [Medline: 10471239] [doi: 10.1016/S1386-5056(99)00018-0]
16. Stumpf SH, Zalunardo RR, Chen RJ. Barriers to telemedicine implementation. Usually it's not technology issues that undermine a project--it's everything else. Healthc Inform 2002 Apr;19(4):45-48. [Medline: 12827765]
17. Lorenzi NM. Beyond the gadgets. BMJ 2004 May 15;328(7449):1146-1147 [FREE Full text] [Medline: 15142892] [doi: 10.1136/bmj.328.7449.1146]
18. Ammenwerth E, Shaw NT. Bad health informatics can kill--is evaluation the answer? Methods Inf Med 2005;44(1):1-3. [Medline: 15778787]
19. Dumay ACM, Freriks G. Quality management issues for medical ICT. Stud Health Technol Inform 2004;103:93-100. [Medline: 15747910]
20. Johnston D, Pan E, Middleton B. The demonstrated value of healthcare IT: decision makers need hard evidence on all potential benefits. Healthc Inform 2003:93-96.
21. Pagliari C, Gregor P, Sullivan F, Sloan D, McGillivray S, Ricketts I, et al. Literature review and conceptual map of the area of e-health. Presentation at: Fourth National SDO Conference; April 27, London, UK. 2005. URL: http://www.lshtm.ac.uk/hsru/sdo/files/adhoc/conference-2005-pagliari.pdf [accessed 2007 May 11] [WebCite Cache ID 5KF5UDLv0]
22. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [Medline: 16702590]
23. Fitzgerald B. An empirical investigation into the adoption of systems development methodologies. Information & Management 1998;34:317-328 [FREE Full text] [doi: 10.1016/S0378-7206(98)00072-X]
24. Khoo B. A Survey of Major Software Design Methodologies. Khoo Homepage. URL: http://userpages.umbc.edu/~khoo/survey2.html [accessed 2007 May 11] [WebCite Cache ID 5OlFR53up]
25. Royce WW. Managing the Development of Large Software Systems. Proceedings of IEEE WESCON. 1970 Aug. p. 1-9 URL: http://www.cs.umd.edu/class/spring2003/cmsc838p/Process/waterfall.pdf [accessed 2007 May 14] [WebCite Cache ID 5KFMYybfW]
26. Boehm BW. A spiral model of software development and enhancement. Computer. 1988. p. 61-72 URL: http://www.sce.carleton.ca/faculty/ajila/4106-5006/Spiral%20Model%20Boehm.pdf [accessed 2007 May 14] [WebCite Cache ID 5KFMOjIVg]
27. Hartson HR, Hix D. Toward empirically derived methodologies and tools for human-computer interface development. Int J Man Mach Stud 1989;31(4):477-494. [doi: 10.1016/0020-7373(89)90005-9]
28. Laplante PA, Neill CJ. ACM. 2004 Feb. (10) URL: http://www.acmqueue.com/modules.php?name=Content&pa=showpage&pid=110 [accessed 2007 May 14] [WebCite Cache ID 5KF351Syn]
29. ; International Organization for Standardization (ISO). ISO 13407:1999 (2004 version). Human-centred design processes for interactive systems. ISO. URL: http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=21197 [accessed 2007 May 14] [WebCite Cache ID 5KF3Nt6hZ]
30. Thursky KA, Mahemoff M. User-centered design techniques for a computerised antibiotic decision support system in an intensive care unit. Int J Med Inform 2006 Sep 1. [Medline: 16950650] [doi: 10.1016/j.ijmedinf.2006.07.011]
31. Waller A, Franklin V, Pagliari C, Greene S. Participatory design of a text message scheduling system to support young people with diabetes. Health Informatics J 2006 Dec;12(4):304-318. [Medline: 17093001] [doi: 10.1177/1460458206070023]
32. Bowman M, Lopez A, Donlon K, Tecuci G. Teaching Intelligent Agents: Software Design Methodology. Crosstalk: J Defense Software Engineering. 2001 Jun. (6) p. 10-14 URL: http://www.stsc.hill.af.mil/crossTalk/2001/06/bowman.html [accessed 2007 May 14] [WebCite Cache ID 5KF3Laozm]
33. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696 [FREE Full text] [Medline: 10987780] [doi: 10.1136/bmj.321.7262.694]
34. Zuber-Skerrit O. Models for action research. In: Pinchen S, Passfield R, editors. Moving On: Creative Applications of Action Learning and Action Research. Queensland, Australia: Action Research, Action Learning and Process Management; 1995:3-29.
35. Deming WE. Out of the Crisis. Cambridge, MA: MIT Press; 2000.
36. How to Improve. Institute for Healthcare Improvement. URL: http://www.ihi.org/IHI/Topics/Improvement/ImprovementMethods/HowToImprove/ [accessed 2007 May 14] [WebCite Cache ID 5KF3SsWNQ]
37. Lewis W. Software Testing and Continuous Quality Improvement. 2nd edition. Boca Raton, FL: CRC Press; 2000.
38. Kushniruk AW, Patel VL, Cimino JJ. Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces. Proc AMIA Annu Fall Symp 1997:218-222. [Medline: 9357620]
39. Jaspers MWM, Steen T. Cognitive engineering in interface design. Stud Health Technol Inform 2002;90:123-127. [Medline: 15460673]
40. Jaspers MWM, Steen T, Van Den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004 Nov;73(11-12):781-795. [Medline: 15491929] [doi: 10.1016/j.ijmedinf.2004.08.003]

41. Friedman CP, Wyatt JC. Evaluation methods in biomedical informatics (Health informatics). New York: Springer-Verlag; 2006.
42. Coiera E, Westbrook J, Wyatt J. The safety and quality of decision support systems. Methods Inf Med 2006;45 Suppl 1(suppl 1):20-25. [Medline: 17051290]
43. ; Department of Trade and Industry. Competing in the Global Economy: The Innovation Challenge. Innovation Report. UK Department of Trade and Industry. London, UK: HMSO; 2003. URL: http://www.dti.gov.uk/files/file12093.pdf [accessed 2007 May 14] [WebCite Cache ID 5KF3dxFye]
44. Montgomery AA, Fahey T, Peters TJ, Macintosh C, Sharp DJ. Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomised controlled trial. BMJ 2000 Mar 11;320(7236):686-690 [FREE Full text] [Medline: 10710578] [doi: 10.1136/bmj.320.7236.686]
45. Machan C, Ammenweth E, Bodner T. Publication bias in medical informatics evaluation literature: recognizing the problem, its impact and the causes. Conference paper 51. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie; September 14, 2006. 2006. URL: http://www.gmds2006.de/Abstracts/129.pdf [accessed 2007 May 14] [WebCite Cache ID 5KF4RF78o]
46. Heathfield H, Pitty D, Hanka R. Evaluating information technology in health care: barriers and challenges. BMJ 1998 Jun 27;316(7149):1959-1961 [FREE Full text] [Medline: 9641938]
47. Gustafson DH, Wyatt JC. Evaluation of ehealth systems and services. BMJ 2004 May 15;328(7449):1150 [FREE Full text] [Medline: 15142895] [doi: 10.1136/bmj.328.7449.1150]
48. Ammenwerth E, De Keizer N. An inventory of evaluation studies of information technology in health care trends in evaluation research 1982-2002. Methods Inf Med 2005;44(1):44-56. [Medline: 15778794]
49. Ewusi-Mensah K. Software Development Failures. Cambridge, MA: MIT Press; 2003.
50. 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]
51. Anderson GF, Frogner BK, Johns RA, Reinhardt UE. Health care spending and use of information technology in OECD countries. Health Aff (Millwood) 2006 May;25(3):819-831. [Medline: 16684749] [doi: 10.1377/hlthaff.25.3.819]
52. Clegg M. Commissioning evaluation: is there an enlightened approach? Proceedings of the UK Evaluation Society Conference; December. 2002. URL: http://www.evaluation.org.uk/conference/Conf%20presentations%202002/Clegg.pdf [accessed 2007 May 14] [WebCite Cache ID 5KF4eFXOd]
53. Pagliari C. Implementing the National Programme for IT: what can we learn from the Scottish experience? Inform Prim Care 2005;13(2):105-111. [Medline: 15992495]
54. ; DG Enterprise and Industry of the European Commission. Supporting the Monitoring and Evaluation of Innovation Programmes. A Study for DG Enterprise and Industry. Final Report. Brussels-Luxembourg: DG Enterprise and Industry of the European Commission; 2006. URL: ftp://ftp.cordis.europa.eu/pub/innovation-policy/studies/smeip_finalreport_master2.pdf [accessed 2007 May 14]

Abbreviations

HSR: health services research

submitted 09.11.06; peer-reviewed by E Ammenwerth, J Aarts; comments to author 14.02.07; revised version received 19.04.07; accepted 04.05.07; published 27.05.07.

Please cite as:
Pagliari C
Design and Evaluation in eHealth: Challenges and Implications for an Interdisciplinary Field
J Med Internet Res 2007;9(2):e15
URL: http://www.jmir.org/2007/2/e15/
doi:10.2196/jmir.9.2.e15
PMID:17537718

Original Paper

A Field Test of a Web-Based Workplace Health Promotion Program to Improve Dietary Practices, Reduce Stress, and Increase Physical Activity: Randomized Controlled Trial

Royer F Cook[1], PhD; Douglas W Billings[1], PhD; Rebekah K Hersch[1], PhD; Anita S Back[1], MS; April Hendrickson[1], MA

ISA Associates, Alexandria, VA, USA

Corresponding Author:
Royer F Cook, PhD
ISA Associates
201 North Union Street
Suite 330
Alexandria, VA 22314
USA
Phone: +1 703 739 0880
Fax: +1 703 739 0462
Email: rcook@isagroup.com

Abstract

Background: Most work sites engage in some form of health promotion programming designed to improve worker health and reduce health care costs. Although these programs have typically been delivered through combinations of seminars and print materials, workplace health promotion programs are increasingly being delivered through the Internet.

Objective: The purpose of this research was to evaluate the effectiveness of a Web-based multimedia health promotion program for the workplace, designed to improve dietary practices, reduce stress, and increase physical activity.

Methods: Using a randomized controlled trial design with pretest-posttest comparisons within each group, 419 employees of a human resources company were randomly assigned to the Web-based condition or to a condition that provided print materials on the same topics. All subjects were assessed at pretest and posttest through an online questionnaire containing multiple measures of health behavior and attitudes. The test period was 3 months. Questionnaire data were analyzed mainly by analysis of covariance and *t* tests.

Results: Retention rates were good for both groups—85% for the Web-based group and 87% for the print group. Subjects using the Web-based program performed significantly better than the print group on Attitudes Toward a Healthful Diet ($F_{1,415}$ = 7.104, P = .008) and Dietary Stage of Change ($F_{1,408}$ = 6.487, P = .01), but there were no significant group differences on the five other dietary measures. Both groups also showed improvement from pretest to posttest on most dietary measures, as indicated by significant *t* tests. Within the Web-based group, dosage analyses showed significant effects of the number of times the subject accessed the program on measures of Dietary Self-Efficacy ($F_{2,203}$ = 5.270, P = .003), Attitudes Toward a Healthful Diet ($F_{2,204}$ = 2.585, P = .045), and Dietary Stage of Change ($F_{2,200}$ = 4.627, P = .005). No significant differences were found between the two groups on measures of stress or physical activity, although *t* tests of pretest-posttest changes indicated that both groups improved on several of these measures. The Web-based group gave significantly higher ratings to the program materials than the print group on all health topics and in their overall evaluation ($F_{1,410}$ = 9.808, P = .002).

Conclusions: The Web-based program was more effective than print materials in producing improvements in the areas of diet and nutrition but was not more effective in reducing stress or increasing physical activity. The higher ratings given to the Web-based program suggest that workers preferred it to the print materials. Both groups showed numerous pretest-posttest improvements in all health topics, although such improvements might be attributable in part to a Hawthorne effect. Results suggest that a multimedia Web-based program can be a promising means of delivering health promotion material to the workforce, particularly in the area of diet and nutrition.

(J Med Internet Res 2007;9(2):e17) doi:10.2196/jmir.9.2.e17

KEYWORDS

Web-based; interventions; stress; diet; activity

Introduction

Workplace health promotion activities and modes of delivery vary widely, from ad hoc events, such as health fairs and provision of print materials, to comprehensive programs involving health risk appraisals, preventive interventions addressing fitness and dietary practices, and intensive disease management programs. A recent comprehensive review of the clinical efficacy and cost-effectiveness of work site comprehensive health promotion and disease management programs concluded that "studies to date indicate positive clinical and cost outcomes" [1]. Although this assessment specifically refers to *comprehensive* programs that focus on risk reduction (particularly for high-risk employees), the generally positive conclusion is congruent with other studies demonstrating effectiveness and cost savings from health promotion and disease management programs at work sites [2,3].

Most of the research conducted to date on work site health has focused on traditional approaches involving in-person (group or individual) interventions, often supplemented with video and print materials. However, computer-based interventions are beginning to emerge, in the workplace as well as in the home, spurred by the tremendous growth in access to the Internet and in the creation of health improvement programs available on the Web. In a review article, Evers points out that The Pew Internet & American Life Project now splits Internet access in the United States into three tiers: those who are truly offline (22% of adults), those with modest connections such as dial-up (40%), and those who are the highly wired broadband elite (33%) [4]. Workplaces are rapidly joining this "elite" group, and the Internet has become increasingly used as a channel for health interventions. Evers notes that despite the familiar caveats about the emerging nature of the data on the effectiveness of Web-based interventions, there is increasing excitement about the potential for Internet technology to facilitate the development of interactive, tailored, multimedia behavior change programs [4]. Of particular note are the results of a study by Wantland and associates, whose meta-analysis of 22 studies found that Web-based health interventions demonstrated improved outcomes over non-Web-based interventions [5]. And in a recent evaluation of a Web-based training program for health promotion practitioners, our research team found the program to be more effective than print materials [6]. Yet, despite the proliferation of Web-based health improvement programs in the workplace, in comparison to traditional modes of delivery there has been relatively little evaluation of Web-based workplace interventions, particularly preventive interventions targeting multiple health behaviors.

The purpose of this study was to test the efficacy of a Web-based health promotion program for the workplace designed to improve participants' dietary practices, increase their level of physical activity, and reduce their stress. The research sought to (1) compare the program's effectiveness with a print-based intervention; (2) assess the program's effectiveness in achieving positive changes in multiple self-report measures of health status, attitudes, and practices; and (3) assess participants' reactions to the Web-based program compared to the print materials.

It should be understood that the purpose of the research was *not* to compare material presented in the Web-based format with the exactly same material presented in print. Instead, the purpose was to test the efficacy of a multimedia Web-based program (created in Macromedia Flash) compared to high-quality commercially available print materials on the same topics (but not necessarily the same content) as a control.

This research represents a continuation of our group's research on the development and testing of video-centered programs designed to improve workforce health and reduce substance abuse [7-11]. As with the video-centered programs, the present program was shaped by a social-cognitive conceptual model based mainly on the work of Bandura [12,13] and emphasizing observational learning, boosting of self-efficacy, and self-regulation. This model was first adapted to workforce health by the authors of a 1990 monograph [14] and was further refined as suggested by the data from subsequent studies (see [9] for the most recent discussion of the model). In addition to social-cognitive theory, the model also draws on the transtheoretical model of Prochaska and associates [15], recognizing that individuals are typically at one of several stages in their readiness to change health behaviors, that interventions must be appropriate to their stage of change, and that moving individuals from one stage to another can be as important as effecting behavior change.

The central hypotheses of the study were as follows:

1. The Web-based program group will show significantly greater improvement in the outcome measures of diet, stress, and physical activity than the print group.
2. The Web-based program group will show significant improvement in the outcome measures from pretest to posttest.
3. The Web-based program will be rated as more informative and appealing than the print materials.

Methods

Design

The study was a randomized controlled trial with pretest-posttest comparisons in each group. Employees in three offices of a human resources company were recruited and randomly assigned to the Web-based program condition or the print condition (receiving print materials covering the same health topics as the Web-based program). It was recognized that the print materials could also be effective instruments of health behavior improvement (unlike a no-treatment control group) and could be a challenge as a control group. Indeed, studies have found print materials to be effective in improving behaviors and outcomes related to walking and other physical activities [16,17], alcohol problems [18], and emotional disorders [19].

However, we believed that the distribution of print materials would be a likely workplace alternative to an online program; therefore, the print group was thought to be an appropriate control group for the study.

The field test was conducted at a major human resources provider with a workforce of approximately 5000 employees. The test was conducted at three of the company's offices located in Atlanta, GA, Minneapolis, MN, and Fountain Valley, CA. The study took place from August 1, 2004 until July 31st, 2005, with recruitment starting in January 2005 and data collection ending in June 2005. Recruitment procedures included an email letter from management describing the purpose and nature of the field test and the program, as well as the incentives for participating (US $50/survey and a US $500 raffle prize). An online flyer was attached to the email, presenting the basic information in a more graphic, colorful fashion. In addition, posters describing the program and the field test were placed throughout the participating offices at the start of participant recruitment. The recruitment period lasted 3 weeks, followed by the pretest of all participants.

The pretest and posttest surveys were online survey assessments containing multiple self-reports of health practice and attitudes (described below). The pretest period lasted 2 weeks. Following this, the Web-based group was provided with codes to access the online program, and the print group was provided with a packet of print materials. The test period lasted 3 months, followed by the posttest survey of all participants.

The study methods and procedures were approved and periodically reviewed by the Institutional Review Board of ISA Associates, which has a Federal Wide Assurance from the Office of Human Research Protections of the National Institutes of Health.

Sample

A total of 480 employees participated in the pretest assessment, and 419 completed the full posttest surveys. 209 were in the Web group, and 210 were in the print group. The characteristics of the participants are presented in Table 1.

Table 1. Demographic characteristics of field test participants

Characteristic	Web Group (N = 209)		Print Group (N = 210)	
	n	**%**	**n**	**%**
Race				
White	163	78.0	177	84.3
African American	18	8.6	11	5.2
Asian	12	5.7	9	4.3
Native Alaskan / Pacific Islander	2	1.0	1	0.5
Other	7	3.3	6	2.9
Ethnicity				
Hispanic	10	4.8	10	4.8
Non-Hispanic	184	88.0	191	91.0
Gender				
Male	53	25.4	62	29.5
Female	155	74.2	147	70.0
Education				
Less than high school	1	0.5	3	1.4
High school	7	3.3	8	3.8
Some college	70	33.5	71	33.8
Bachelor's degree	79	37.8	78	37.1
Postgraduate degree	50	23.9	48	22.9
Marital Status				
Single	66	31.6	60	28.6
Married	125	59.8	28	61.0
Domestic partnership	10	4.8	9	4.3
Divorced	3	1.4	8	3.8
Widowed	2	1.0	1	0.5
Other	0	0.0	2	1.0
Annual Income (US $)				
20000-29999	0	0.0	3	1.4
30000-39999	13	6.2	18	8.6
40000-49999	32	15.3	19	9.0
Over 50000	143	68.4	154	73.3
Age (years), mean (range)	41.99 (24-64)		42.03 (22-66)	

Measures

The measures contained in the online health surveys are listed below. The majority of these measures have been used in previously published studies by this research team and have shown evidence of reliability and validity in a variety of workplace settings and populations [7-11].

- Attitudes Toward a Healthful Diet: An 18-item scale assessing one's attitudes toward healthful eating practices (alpha = .70). Higher score is better.

- Eating Practices: A 10-item scale assessing the frequency with which one engages in healthful eating practices (alpha = .63). Higher score is better.
- Motivation to Improve Diet: One item asking how important a healthy diet is to the respondent. Higher score is better.
- Dietary Behavioral Intentions: A 5-item scale assessing one's intentions to eat a healthful diet (alpha = .86). Higher score is better.
- Dietary Self-Efficacy: A 6-item scale assessing one's confidence in being able to eat a healthful diet (alpha = .89). Higher score is better.
- Dietary Stage of Change: A 4-item scale assessing one's stage of change in adopting a healthful diet (alpha = .74). Lower score is better.
- Weight Stage of Change: A 4-item algorithm designed to classify one's stage of change in losing weight or avoiding gaining weight [20]. Higher score is better.
- Perceived Stress: A 5-item scale assessing one's level of perceived stress (alpha = .89). Lower score is better.
- Symptoms of Distress: A 15-item scale assessing one's level of physical (alpha = .71) and emotional (alpha = .88) symptoms of stress. Lower score is better.
- Stress Stage of Change: One item assessing one's stage of change in attempting to reduce stress. Lower score is better.
- Brief COPE: A 26-item measure, consisting of 13 two-item subscales, assessing one's skills in coping with stress (alpha range: .47-.88) [21].
- Godin Leisure-Time Exercise Questionnaire: A 3-item scale assessing the frequency of physical activity in the past week (alpha = .78) [22]. Higher score is better.
- Godin Sweat Score: One item assessing one's frequency of engaging in strenuous activity. Lower score is better.
- Exercise Stage of Change: One item assessing one's stage of change in adhering to regular physical activity. Lower score is better
- Exercise Motivation: One item asking how important physical activity is to the respondent. Higher score is better.
- Exercise Behavioral Intentions: One item assessing one's intentions to engage in regular physical activity. Higher score is better.
- Exercise Self-Efficacy: One item assessing one's confidence in engaging in regular physical activity. Higher score is better.
- Weight: One item asking the respondent's current weight.
- User Evaluations of Program (posttest-only): Five items in each of four health topics (stress, weight, diet/nutrition, and substance use) assessing the extent to which the materials (Web-based or print) provided information, helpful tips, motivation, good examples, and encouragement to examine practices. Lower score is better.
- Time and Frequency of Access (posttest Web group only): Respondent was asked to estimate frequency of and length of time (in minutes) spent accessing specific segments of the Web-based program.
- Ratings of Program Functions (posttest Web group only): A 10-item scale asking the respondent to rate the functionality (eg, ease of navigation, clarity of layout) of the program. Lower score is better.

A copy of all health outcome measures is available upon request.

Equivalence of Groups

To determine the degree to which the randomization produced like samples, the experimental and control groups were compared on demographics and selected pretest dependent measures. To assess whether the experimental and control groups of the main study sample (those who completed both pretest and posttest) differed on selected dependent measures at pretest, analysis of variance (ANOVA) contrasted the two groups at pretest on five key dependent measures—Attitudes Toward a Healthful Diet, Eating Practices, Symptoms of Distress, Godin Sweat Score [22], and Weight. All comparisons were nonsignificant, indicating that the two groups were equivalent at pretest on the central outcomes of interest. Chi-square analyses contrasted the two groups on demographics at pretest—race, ethnicity, gender, education, marital status, and income—and found no significant differences, indicating that the two groups were demographically equivalent. The results of these analyses indicate that the randomization process was successful in producing equivalent groups.

ANOVA contrasting the pretest scores of the dropouts with those who completed both pretest and posttest on the same five dependent measures (Attitudes Toward a Healthful Diet, Eating Practices, Symptoms of Distress, Godin Sweat Score, and Weight) found no significant differences between the two groups, indicating that there was not differential attrition as a function of pretest position on key outcomes. Chi-square analyses contrasted the demographics (race, ethnicity, gender, education, marital status, and income) of the dropouts with those who completed both surveys and found no differences between the two groups, indicating that the demographic composition of the dropouts did not differ significantly from that of the participants who completed both surveys.

Interventions

The Web-based program, Health Connection, is a comprehensive multimedia health promotion program offering substantial information and guidance on the major health promotion and wellness topics of stress management, nutrition/weight management, and fitness/physical activity. Conceptually rooted in accepted models of health behavior change as described above [12-13,15], the program was intended to be more than an information resource: it was designed to improve the health practices and attitudes of working adults in the three health topic areas. Many of the concepts and much of the content of the program were drawn from previous video-centered programs developed and tested by our group in a series of workplace-based studies [7-11].

The program was developed by the ISA Associates team over a 2-year period through multiple cycles of development and testing, beginning with focus group assessments of prototype content and culminating in the workplace-based field test. From the outset, the program was designed specifically for the broadband environment, anticipating that most workplaces would have DSL or higher Internet connections by 2005. Accordingly, the program, constructed in Macromedia Flash, is highly interactive with ample graphics, audio, and video (the entire program is audio-narrated). Many of these main elements are congruent with health behavior change theory and principles

(eg, providing opportunities for observational learning, building self-efficacy, and self-tailoring of content and sequence). Sample screenshots are displayed in Figures 1a and 1b and Multimedia Appendix 1, and an outline of the program content is available in Multimedia Appendix 2.

The print materials consisted of five commercially available booklets:

1. "Low-Fat Eating" (15 pages)
2. "Getting Started with Weight Management" (15 pages)
3. "Stress Management: A Personal Action Guide" (15 pages)
4. "Fitness: The High Performance Lifestyle" (17 pages)
5. "Alcohol, Drugs, and a Healthy Lifestyle: What's the Connection" (11 pages)

All booklets were colorful and included graphics and various tracking forms and logs. Outlines of the content of each booklet are presented in Multimedia Appendix 3.

Figure 1a. Health Connection screenshots

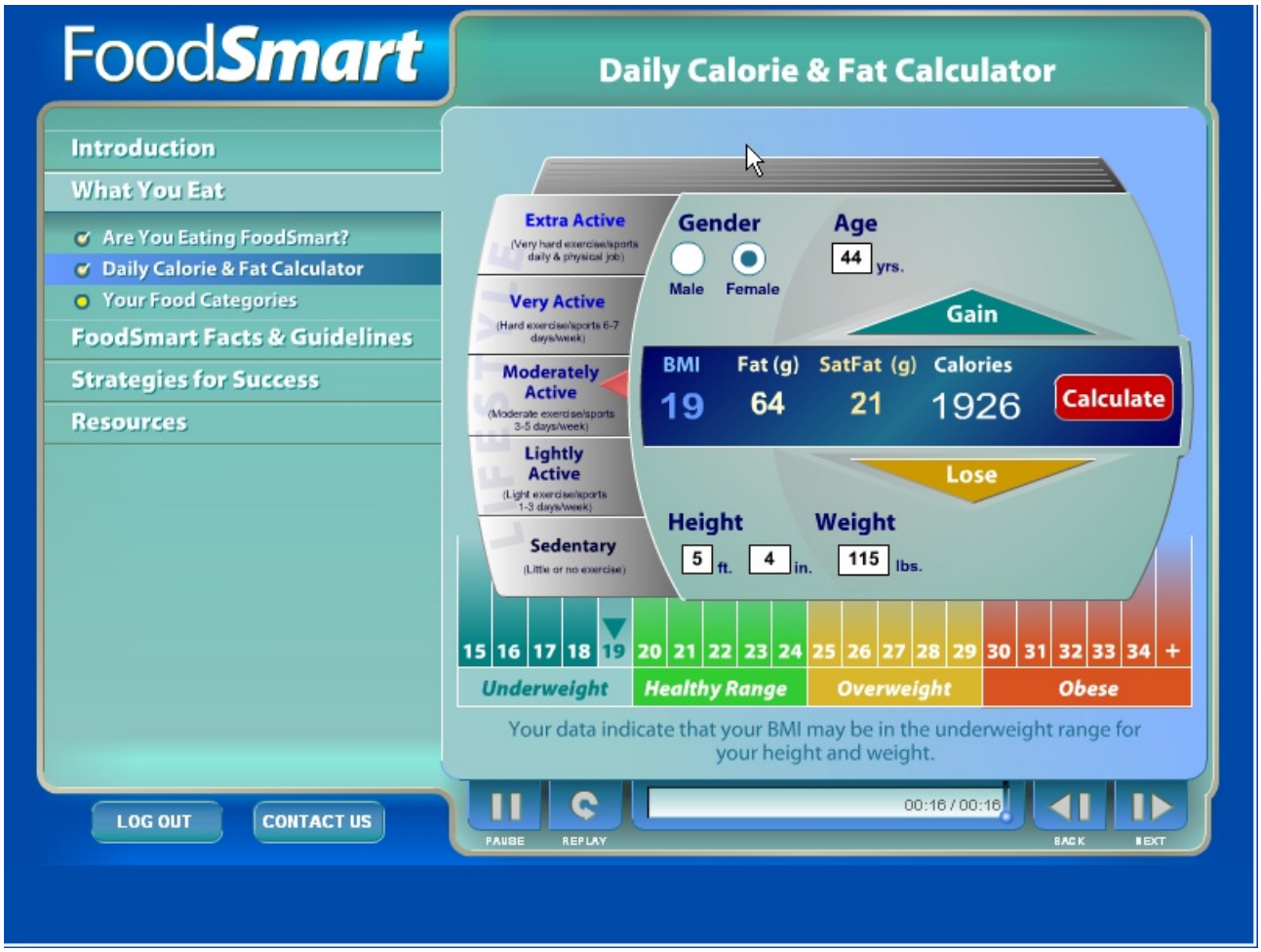
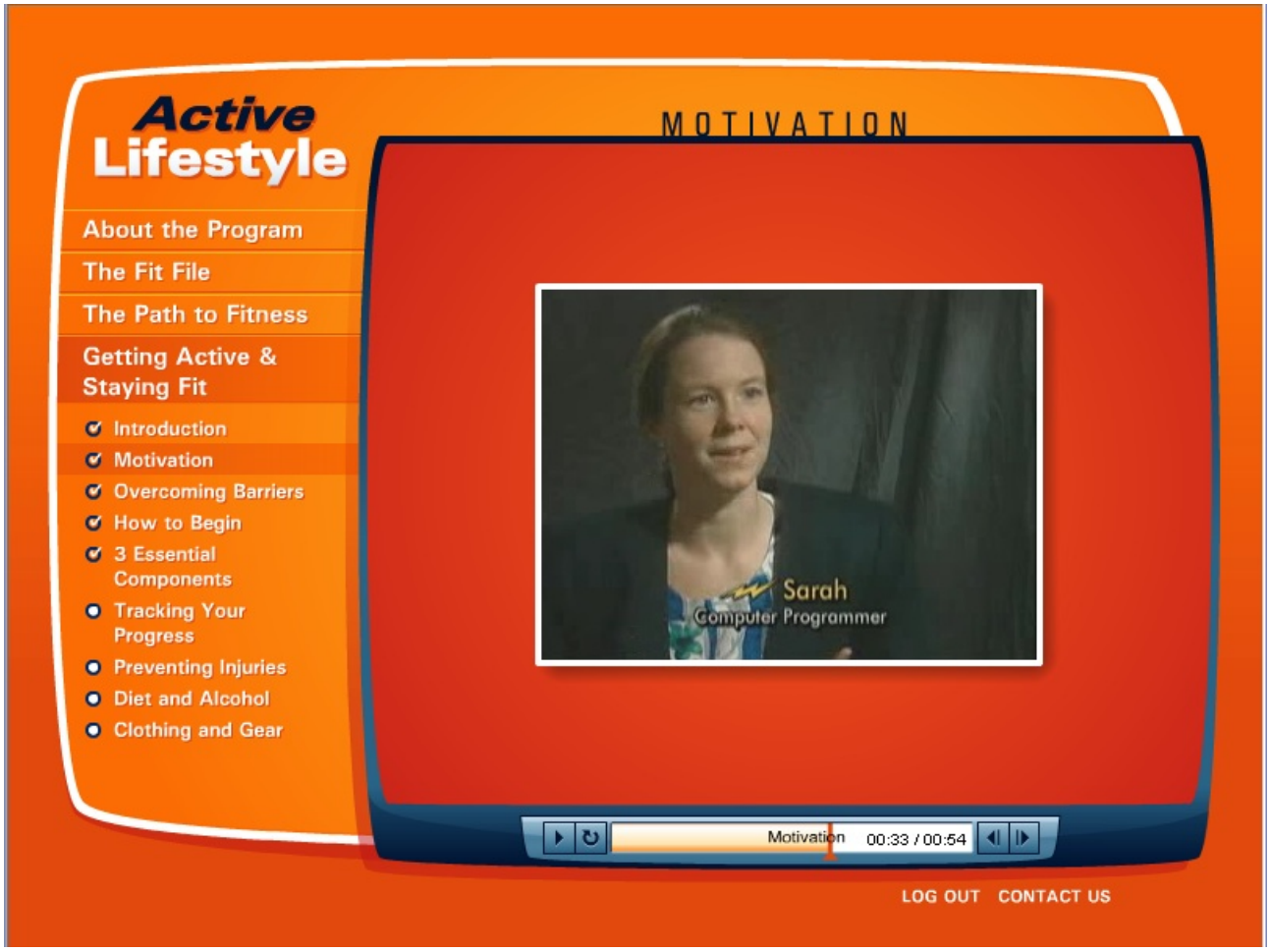


Figure 1b. Health Connection screenshots



Analysis

Four major types of analyses were conducted on the health outcomes survey data: (1) analysis of covariance (ANCOVA), contrasting posttest measures of the Web and print groups with pretest measures as the covariate; (2) *t* tests of pretest-posttest changes within group; (3) dosage analyses using ANOVA to test for Web program exposure effects on dependent measures; and (4) ANOVA to compare user evaluations of the program (ratings of Web program vs print materials).

Results

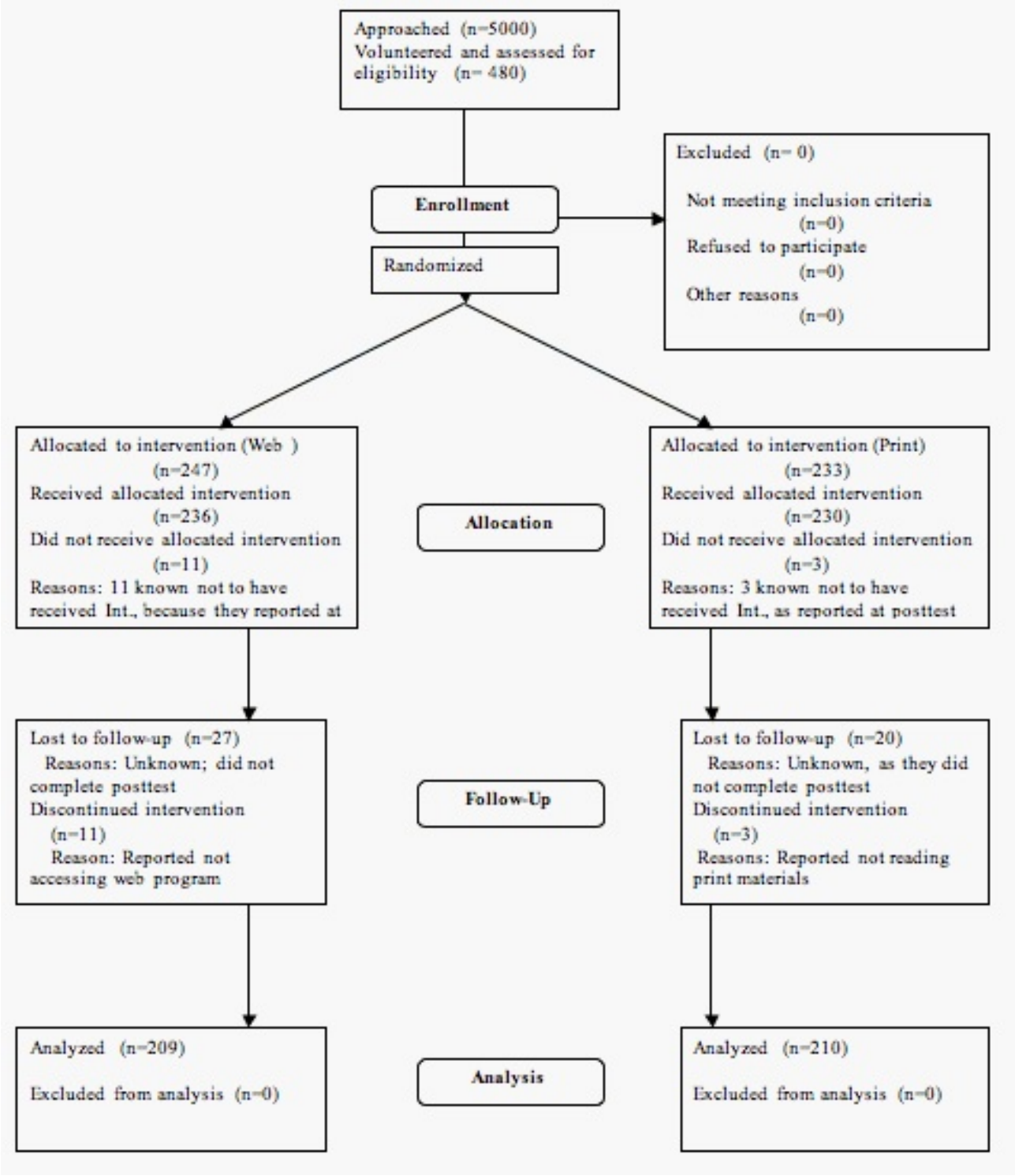
Participation and Retention

There were approximately 5000 employees in the company, virtually all of whom were eligible to participate in the program. A total of 480 employees signed up for the program and participated in the pretest survey, roughly a 10% participation rate.

There was a moderate amount of attrition from pretest to posttest, as 61 subjects — 38 from the Web group (85% retention rate) and 23 from the print group (87% retention rate) — who completed the pretest were not included in the final analysis, either because they did not complete the posttest or because they did not meet the criteria for inclusion. In the Web group, 27 subjects did not complete the posttest, and 11 subjects were excluded because they reported that they did not access the program. In the print group, 20 subjects did not complete the posttest, and 3 subjects claimed never to have received the print materials. The flow of participants through the study is shown in Figure 2.

For a more detailed analysis and discussion of the rates of participation and retention (along with frequency of Web program access), see Multimedia Appendix 4 for a PowerPoint presentation by Cook and associates.

Figure 2. CONSORT Flowchart, displaying the flow of participants from enrollment through analysis.



Eating Practices and Attitudes

Table 2 presents the results of the ANCOVA to test differential improvement on several measures of eating practices and attitudes. The Web group showed significantly greater improvement than the print group on measures of Attitudes Toward a Healthful Diet and Dietary Stage of Change. There were no other differences in the ANCOVA results between the two groups on these measures.

Table 2. Eating practices and attitudes

Measure	Web Posttest Mean (SD)	Print Posttest Mean (SD)	F (Degrees of Freedom)	*P*
Attitudes Toward a Healthful Diet*	3.835 (0.395)	3.711 (0.434)	7.104 (1,415)	.008
Eating Patterns*	2.695 (0.488)	2.641 (0.474)	0.004 (1,416)	.95
Dietary Motivation*	4.24 (0.786)	4.18 (0.850)	0.012 (1,416)	.91
Dietary Intentions*	4.126 (0.774)	4.189 (0.731)	0.198 (1,414)	.66
Dietary Self-Efficacy*	3.840 (0.759)	3.771 (0.784)	0.092 (1,414)	.76
Dietary Stage of Change†	1.902 (0.902)	2.149 (0.921)	6.487 (1,408)	.01
Weight Stage of Change*	3.00 (0.909)	3.00 (0.712)	0.286 (1,403)	.59

*Higher score is better.

†Lower score is better.

Table 3 shows the results of tests of pretest-posttest changes in dietary outcomes for the Web group, showing that the Web group produced significant pretest-posttest improvements on all seven outcome measures. However, the print group showed similar significant improvements, with the exception of a nonsignificant change in Weight Stage of Change.

Table 3. Pretest-posttest changes: eating practices and attitudes

Measure	Pretest Mean (SD)	Posttest Mean (SD)	*t*	*P*
Attitudes Toward a Healthful Diet*				
Web	3.6358 (0.4309)	3.8354 (0.3950)	−7.673	< .001
Print	3.5698 (0.4586)	3.7113 (0.4344)	−5.051	< .001
Eating Patterns*				
Web	2.5929 (0.50295)	2.6949 (0.4884)	−4.717	< .001
Print	2.5256 (0.4669)	2.6408 (0.4743)	−5.174	< .001
Dietary Motivation*				
Web	4.24 (0.786)	4.13 (0.783)	2.506	.01
Print	4.18 (0.850)	4.04 (0.797)	2.771	.01
Dietary Intentions*				
Web	3.9116 (0.8713)	4.1256 (0.7736)	−4.147	< .001
Print	3.984 (0.7979)	4.1895 (0.7313)	−3.972	< .001
Dietary Self-Efficacy*				
Web	3.5533 (0.8207)	3.8403 (0.7587)	−5.494	< .001
Print	3.4562 (0.8503)	3.7711 (0.7844)	−5.978	< .001
Dietary Stage of Change†				
Web	2.306 (1.061)	1.902 (0.902)	6.167	< .001
Print	2.4199 (0.9989)	2.1486 (0.9214)	4.385	< .001
Weight Stage of Change*				
Web	2.84 (0.880)	3.00 (0.909)	−2.747	.01
Print	2.92 (0.735)	3.00 (0.712)	−1.694	.09

*Higher score is better.

†Lower score is better.

Weight

The ANCOVA analyses for reported weight indicated that there was no significant differential change in weight between the two groups. However, both groups achieved a significant, though small, reduction in reported weight from pretest to posttest. The Web group lost an average of 0.57 kg (t = 2.09, P = .04), and the print group lost an average of 0.96 kg (t = 2.33, P = .02).

Stress Management

The ANCOVA to test differential change between the Web group and the print group detected no differences between the two groups on the three measures of stress. In addition, none of the 13 subscales of Coping Skills (2-item subscales, scored individually, equaling 26 items) yielded differences between the groups.

Table 4 displays the results of tests of pretest-posttest change in stress outcomes, showing significant improvement by the Web group on Symptoms of Distress; however, the print group showed significant pretest-posttest improvement on all three stress outcome measures.

Table 4. Pretest-posttest changes: stress management

Measure*	Pretest Mean (SD)	Posttest Mean (SD)	t	P
Perceived Stress				
Web	14.21 (4.936)	13.6746 (4.8773)	1.780	.08
Print	15.05 (4.645)	14.2714 (4.545)	2.862	.01
Stress Stage of Change				
Web	2.40 (1.555)	2.30 (1.504)	0.915	.36
Print	2.53 (1.535)	2.15 (1.417)	3.716	< .001
Symptoms of Distress				
Web	30.29 (8.22)	29.14 (8.386)	2.474	.01
Print	31.62 (7.922)	30.03 (7.659)	3.813	< .001

*Lower score is better.

Physical Activity

The ANCOVA to test differential change between the Web group and the print group detected no differences between the two groups on the five measures of physical activity. As shown in Table 5, the tests of pretest-posttest changes within groups showed that both groups achieved significant improvement on the Godin Sweat Score [22] and the Activity Stage of Change measure, while the print group also showed significant improvement on the Activity Confidence measure (although Web group came close).

Table 5. Pretest-posttest changes: physical activity

Measure	Pretest Mean (SD)	Posttest Mean (SD)	*t*	*P*
Godin Leisure Score*				
Web	46.55 (55.866)	57.35 (121.588)	−1.688	.09
Print	42.81 (62.115)	46.53 (35.916)	−0.802	.42
Godin Sweat Score†				
Web	2.08 (0.751)	1.91 (0.646)	3.454	.001
Print	2.17 (0.784)	1.94 (0.705)	4.546	< .001
Activity Stage of Change†				
Web	2.26 (1.286)	1.97 (1.191)	3.682	< .001
Print	2.50 (1.279)	2.13 (1.192)	4.556	< .001
Activity Motivation*				
Web	4.12 (0.953)	4.06 (1.052)	1.051	.29
Print	4.14 (0.898)	4.11 (1.098)	0.573	.57
Activity Intentions*				
Web	4.53 (0.791)	4.60 (0.694)	−1.414	.16
Print	4.48 (0.816)	4.56 (0.778)	−1.490	.14
Activity Confidence*				
Web	3.80 (1.194)	3.94 (1.020)	−1.896	.06
Print	3.51 (1.217)	3.81 (1.131)	−4.068	< .001

*Higher score is better.

†Lower score is better.

Dosage Analysis

In order to assess the extent to which the Web program effects were a function of the number of times the user accessed a particular program module (stress management, physical activity, and nutrition/weight control), a series of dosage analyses were conducted on the outcome measures. The Web group participants were classified into three groups: those who reported that they never accessed the particular module, those who reported accessing it once, and those who reported accessing it more than once. The analyses tested for (1) linear trends by group and (2) multiple contrasts by group for all outcome measures. The multiple contrasts also included the print group as a fourth group.

The results of the trend analysis on dietary measures, shown in Table 6, indicate that there were significant linear effects of the Web-based nutrition/weight control module on three of the seven dietary measures: Self-Efficacy, Attitudes Toward a Healthful Diet, and Dietary Stage of Change.

Table 6. Nutrition/weight control module effects by dosage: trend analyses

	Mean Score by Number of Times Accessed (No.)			Overall		Linear
Measure	Never	Once	More Than Once	F (Degrees of Freedom)	*P*	*P*
Dietary Self-Efficacy	3.579 (28)	3.781 (93)	3.989 (86)	5.270 (2,203)	.006	.003
Attitudes Towards a Healthful Diet	3.751 (28)	3.809 (93)	3.890 (87)	2.585 (2,204)	.08	.04
Eating Patterns	2.678 (28)	2.695 (94)	2.700 (87)	.059 (2,205)	.94	.73
Dietary Motivation	4.090 (28)	4.229 (94)	4.310 (87)	1.473 (2,205)	.23	.09
Dietary Intentions	4.110 (27)	4.075 (93)	4.184 (87)	.694 (2,203)	.50	.59
Dietary Stage of Change	2.199 (28)	1.961 (92)	1.738 (84)	4.627 (2,200)	.01	.01
Weight Stage of Change	3.118 (27)	2.894 (92)	3.079 (83)	1.682 (2,198)	.19	.82

The analysis of contrasts of the print group (who never accessed the nutrition/weight control module) with the three different levels of access of the nutrition/weight control module in the Web group also produced significant dosage effects for the same three dietary measures. On all three measures, the significant contrast occurred in the print group versus the Web group that accessed the module more than once (Table 7).

Table 7. Nutrition/weight control module effects by dosage: multiple contrasts

	Contrasts of the Print Group vs Levels of Web Access, *P* values			Overall Contrast	
Measure	Never	Once	More Than Once	F (Degrees of Freedom)	*P*
Dietary Self-Efficacy	.06	.61	.04	3.422 (3,412)	.02
Attitudes Towards a Healthful Diet	.96	.14	.001	3.937 (3,413)	.01
Eating Patterns	.80	.95	.86	0.039 (3,414)	.99
Dietary Motivation	.25	.92	.38	0.883 (3,414)	.45
Dietary Intentions	.74	.31	.69	0.542 (3,412)	.65
Dietary Stage of Change	.48	.17	.00	5.218 (3,406)	.002
Weight Stage of Change	.31	.42	.18	1.457 (3,401)	.23

The findings from the dosage analysis are similar to those for the two-group Web versus print ANCOVA, except that the dosage analysis—both trend and multiple contrasts—revealed significant effects of the Web program on the measure of Dietary Self-Efficacy as well as Attitudes Toward a Healthful Diet, and Dietary Stage of Change.

The dosage analyses conducted for the other two modules (physical activity and stress management) found no significant dosage effects for either.

Program Evaluations

At posttest, all subjects rated the program materials, either the Web-based program or the print materials, using a 5-point scale. In each of three topic areas—stress, diet, and weight management—subjects rated the materials on the degree to which they (1) “provided a wealth of information,” (2) “gave me helpful tips,” (3) “motivated me,” (4) provided good examples,” and (5) “encouraged me to examine [the topic].” Within each topic area, the ratings were averaged and ANOVA was used to test differences between groups on these average ratings. As shown in the first three rows of Table 8, the Web group ratings significantly exceeded the print group in all three topic areas.

Subjects also rated the extent which the materials (Web-based or print) were “engaging and appealing,” and “easy to access and understand.” The last three rows of Table 8 show that the ratings for the Web-based program significantly exceeded the print materials for the first measure and in the average of the two ratings (overall). Although the Web-based program received higher ratings for the second measure, the difference was not statistically significant.

Table 8. Program evaluations

Segment/Topic*	Web Posttest Mean (SD)	Print Posttest Mean (SD)	F (Degrees of Freedom)	*P*
Stress management	2.00 (0.73)	2.15 (0.61)	4.748 (1,395)	.03
Diet/nutrition	1.86 (0.67)	2.08 (0.67)	11.380 (1,398)	.001
Weight management	1.99 (.073)	2.16 (0.69)	5.721 (1,392)	.02
Engaging/appealing	1.79 (0.86)	2.09 (0.87)	12.156 (1,410)	.001
Accessible/ understandable	1.50 (0.55)	1.60 (0.50)	3.233 (1,409)	.07
Overall evaluation	1.65 (0.64)	1.84 (0.60)	9.808 (1,410)	.002

*Lower score is better.

Discussion

Review and Interpretation of Findings

The direct comparisons of the Web group and print group showed that the Web-based program performed significantly better than the print materials on measures of Attitudes Toward a Healthful Diet and Dietary Stage of Change, and the dosage analysis indicated significant effects of the Web-based nutrition/weight management segment on Dietary Self-Efficacy as well. In addition, the Web-based program performed significantly better than the print materials on virtually all dimensions of program evaluation. However, on outcome measures of stress and physical activity, there were no significant differences between those who received the Web-based program and those who received the print materials.

Of particular interest was the finding that both the Web group and the print group showed significant improvements on multiple measures in each topic area (diet, stress, etc) from pretest to posttest, suggesting that either (1) both interventions were effective in improving health practices and attitudes or (2) the improvements were the spurious result of some other force or event, perhaps an overall social desirability effect, a Hawthorne effect, or a function of most subjects in both groups coming to the field test with substantial motivations to improve their health. Because it was originally thought that the print materials would form a relatively weak intervention compared to the Web program, a no-treatment control was not included in the design; thus, it is difficult to determine which interpretation is the correct one. It is possible that the improvements occurred as a result of a secular or environmental change affecting both groups, such as an easing of the workload across the workforce (unlikely) or the implementation of another health promotion program in all three offices (this was not the case). It is also possible that both groups exhibited a quasi-placebo effect: wanting to improve their health practices and given some impetus to do so (the surveys and programs), they reported improvements across most of the measures, but the improvements were not a function of the interventions. However, the fact that the improvements occurred consistently in the predicted direction across the majority of these measures, including reductions in reported weight from pretest to posttest, argues against this interpretation.

Perhaps the most plausible explanation is that most of the participants in both groups entered the study with considerable motivation to make improvements in their health practices and attitudes, and both the Web-based program and the print materials contained sufficient information and guidance to help participants make the desired improvements. Indeed, for both groups, the pretest motivation items for improving dietary practices and for increasing activity exceeded 4.0 on a 5-point scale, indicating that both groups were indeed motivated to make positive changes in diet and activity upon entering the field test. Although the print group was originally conceived as a control group, it seems likely that the scope and quality of the print materials—five colorful booklets that included tracking forms and logs, totaling 73 pages—were such that the print materials were effective interventions in themselves and constituted a substantial rival to the Web program. As noted earlier, the effectiveness of print materials for changing health behaviors and outcomes is supported in the literature [16-19]. In comparison to the print materials, the Web program contained more total content and was presented in a multimedia format; hence, the Web program received higher ratings for being appealing and motivating, providing a wealth of information, and for other evaluation measures than the print materials. Indeed, the superiority of the Web program over the print materials in the participant ratings was striking: across all topic areas, the Web group ratings significantly exceeded the print ratings on 15 items; the print group ratings exceeded the Web group on none. The Web group program consistently garnered higher ratings for “wealth of information,” “helpful tips,” “motivated me,” “provided good examples,” and “encouraged me to examine [the topic].”

On the other hand, the print materials offered other advantages: users had all the material in hand from the start, with no need to access the Web or navigate the program, and they could carry one or more of the booklets virtually anywhere they went. The apparent efficacy of the print materials suggests that increasing the number of printable downloads in the Web program would improve its effectiveness.

The findings from the dosage analysis further support the view that the effects of the Web program were real and not a result of social desirability or other nonprogram effects, at least in the dietary topic area. On multiple measures, the Web program effects were a positive linear function of the number of times users accessed the program: as exposure to the program increased, so did the size of the improvements. As noted in the presentation by Cook et al (see Multimedia Appendix 4), the frequency with which users accessed the Web-based program was lower than anticipated and was especially low for the modules on stress and physical activity, a finding which suggests, along with the dosage analysis, that frequency of access might be related to program effects. Although the causal direction is unclear, taken together these findings suggest that procedures designed to increase the frequency of user access (eg, tying incentives to frequency of access) might improve both adherence and efficacy.

Related Studies

The results of this study are generally congruent with findings from other evaluations of Web-based health interventions, although this Web-based program addressed a wider variety of health topics than most previous interventions, which have typically focused on a single health topic. The meta-analysis of Web-based versus non-Web-based interventions conducted by Wantland and associates concluded that effect size comparisons across the 22 studies showed improvement in knowledge and/or behavior change health outcomes for individuals using Web-based interventions [5]. However, a closer examination of the three studies in the meta-analysis that focused on nutrition/weight management and physical activity reveals that the results for these particular studies were less uniformly positive than the overall conclusions seemed to indicate. The assessment of a Web-based intervention to facilitate weight loss found a modest effect size of .15, and the authors concluded that the Web-based intervention was not as effective as in-person support [23]. A randomized control trial of the efficacy of a Web-based tailored nutrition education program found the Web-based group to be significantly better than the control group on posttest measures of nutrition awareness and intentions to change, concluding that Web-based interventions can lead to changes in *determinants* of behavior [24]. A study comparing Web-based and print versions of a work site physical activity program showed no significant increase in physical activity in either group, although the Web group showed a significant decrease in time reported sitting [25].

However, although the studies of Web-based weight management programs reviewed by Wantland et al did not generate evidence of strong effects, recent evaluations of other Web-based weight management programs [26-28] have produced positive findings, indicating that the positive results

on dietary measures found in this study are congruent with other, perhaps similar Web-based programs. In particular, with its emphasis on healthy eating, the relationship between food consumption and energy expenditure, tips on overcoming barriers, and so on, the Balance program studied by Rothert and associates [28] appears to be quite similar to the Health Connection program, although more structured and focused specifically on weight loss among those with a body mass index above 27. Their program showed substantial effects in the dietary area, as participants in the Balance program lost significantly more weight than participants in an information-only condition, and more Balance participants than information-only participants reported that the program was relevant, helpful, and easy to understand. However, attrition in the Rothert study was much greater than in the present study, as only 30% and 20% of the baseline sample responded to the follow-up surveys at 3 and 6 months, respectively—a significant limitation of the study, as noted by the authors. The much higher retention rate in the present study was probably due in part to the monetary incentives tied to survey responses, while the Rothert study offered no incentives.

In some contrast to the relatively positive findings from this study and others on Web-based dietary and weight management programs, studies of Web-based programs aimed at stress management and increasing physical activity are less numerous and, as indicated by the Wantland review [5], somewhat less efficacious. The relative paucity of Web-based programs aimed at stress and activity compared to weight management is doubtless a reflection of the interests of the general population and the increasing importance placed on weight control by medical professionals. However, the lesser degree of efficacy found in this study as well as others on levels of stress and activity suggests the possibility that these topics might be less suited to Web-based approaches than programs aimed at weight control, or that Web-based techniques in the areas of stress and activity programs are less advanced than in the dietary area.

It should be emphasized that Web-based programs vary widely not only in content but also in the degree to which they are science-based (both scientifically accurate and rooted in accepted theories of health behavior change), media-rich, and interactive. Web-based programs containing questionable information or lacking accepted theoretical foundations are unlikely to be effective [28,29].

There are substantial differences in the level of effort and costs of implementing and distributing Web- and print-based approaches. Providing access to the Web program required only a single email message sent instantaneously to all participants, whereas the print materials had to be mailed individually (through company mail) to all participants located in three different offices across the United States. Providing a Web-based program to users is also probably less costly than providing print materials, except perhaps for very small workforces. Indeed, once built, Web-based programs can be delivered to millions of users in a very cost-effective manner, reaching audience sizes unattainable by the traditional workplace health promotion programs.

Limitations

Limitations of the study include the following: (1) all dependent measures were self-reports; (2) the test period was limited to three months; and (3) a no-treatment control condition was not included. Future research on the efficacy of Web-based health promotion interventions should include the collection of physical/biological measures and health care utilization data, should assess participants for longer periods or time, and should include a no-treatment control group.

Acknowledgments

This study was supported by Contract #N44DA-3-5517 from the US National Institutes of Health.

Conflicts of Interest

The Web-based Health Connection program is owned by the Center for Workforce Health, LLC, in which two of the authors (Drs. Cook and Hersch) have a financial interest.

Multimedia Appendix 1

Selected screenshots of the Health Connection program (pdf) [PDF file (Adobe Acrobat), 748 KB - jmir_v9i2e17_app1.pdf]

Multimedia Appendix 2

Outline of Health Connection Web program content (pdf) [PDF file (Adobe Acrobat), 125 KB - jmir_v9i2e17_app2.pdf]

Multimedia Appendix 3

Detailed outline of print materials reviewed by the control group (pdf) [PDF file (Adobe Acrobat), 103 KB - jmir_v9i2e17_app3.pdf]

Multimedia Appendix 4

Presentation on adherence in Internet interventions aimed at diet, exercise, and stress presented at the 11th World Congress on Internet in Medicine (http://www.mednetcongress.org) (ppt) [PPT file (MS Powerpoint), 152 KB - jmir_v9i2e17_app4.ppt]

References

1. Pelletier KR. A review and analysis of the clinical and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: update VI 2000-2004. J Occup Environ Med 2005 Oct;47(10):1051-1058. [Medline: 16217246] [doi: 10.1097/01.jom.0000174303.85442.bf]
2. Ozminkowski RJ, Ling D, Goetzel RZ, Bruno JA, Rutter KR, Isaac F, et al. Long-term impact of Johnson & Johnson's Health & Wellness Program on health care utilization and expenditures. J Occup Environ Med 2002 Jan;44(1):21-29. [Medline: 11802462] [doi: 10.1097/00043764-200201000-00005]
3. Goetzel RZ, Ozminkowski RJ, Bruno JA, Rutter KR, Isaac F, Wang S. The long-term impact of Johnson & Johnson's Health & Wellness Program on employee health risks. J Occup Environ Med 2002 May;44(5):417-424. [Medline: 12024687] [doi: 10.1097/00043764-200205000-00010]
4. Evers KE. eHealth promotion: the use of the Internet for health promotion. Am J Health Promot 2006;20(4):suppl 1-s7, iii. [Medline: 16555803]
5. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, Mcghee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [Medline: 15631964] [doi: 10.2196/jmir.6.4.e40]
6. McPherson TL, Cook RF, Back AS, Hersch RK, Hendrickson A. A field test of a Web-based substance abuse prevention training program for health promotion professionals. Am J Health Promot 2006;20(6):396-400. [Medline: 16871819]
7. Cook RF, Back A, Trudeau J. Substance abuse prevention in the workplace: recent findings and an expanded conceptual model. J Prim Prev 1996;6(3):319-339. [doi: 10.1007/BF02407428]
8. Cook RF, Back AS, Trudeau J. Preventing alcohol use problems among blue-collar workers: a field test of the Working People program. Subst Use Misuse 1996 Feb;31(3):255-275. [Medline: 8834262]
9. Cook RF, Back AS, Trudeau JV, McPherson TL. Integrating substance abuse prevention into health promotion programs in the workplace: a social cognitive intervention targeting the mainstream user. In: Bennett JB, Lehman WEK, editors. Beyond Drug Testing: Innovative Approaches to Dealing with Employee Substance Abuse. Washington, DC: APA; 2003:97-133.
10. Cook RF, Hersch RK, Back AS, McPherson TL. The prevention of substance abuse among construction workers: a field test of a social cognitive program. J Prim Prev 2004;25(3):337-358. [doi: 10.1023/B:JOPP.0000048025.11036.32]
11. Deitz D, Cook R, Hersch R. Workplace health promotion and utilization of health services: follow-up data findings. J Behav Health Serv Res 2005;32(3):306-319. [Medline: 16010186] [doi: 10.1007/BF02291830]
12. Bandura A. Social Learning Theory. Englewood Cliffs, NJ: Prentice Hall; 1977.
13. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall; 1986.
14. Cook RF, Youngblood A. Preventing substance abuse as an integral part of worksite health promotion. Occup Med 1990;5(4):725-738. [Medline: 2237701]
15. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [Medline: 1329589] [doi: 10.1037/0003-066X.47.9.1102]
16. Humpel N, Marshall AL, Iverson D, Leslie E, Owen N. Trial of print and telephone delivered interventions to influence walking. Prev Med 2004 Sep;39(3):635-641. [Medline: 15313106] [doi: 10.1016/j.ypmed.2004.02.032]
17. Marshall AL, Bauman AE, Owen N, Booth ML, Crawford D, Marcus BH. Population-based randomized controlled trial of a stage-targeted physical activity intervention. Ann Behav Med 2003;25(3):194-202. [Medline: 12763714] [doi: 10.1207/S15324796ABM2503_05]
18. 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]
19. Den Boer PC, 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] [doi: 10.1017/S003329170300179X]
20. Weight: Stages of Change - Short Form. Cancer Prevention Research Center. URL: http://www.uri.edu/research/cprc/Measures/Weight01.htm [accessed 2007 May 24] [WebCite Cache ID 5HcP5MOQz]
21. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [Medline: 16250744] [doi: 10.1207/s15327558ijbm0401_6]
22. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 1985 Sep;10(3):141-146. [Medline: 4053261]
23. Harvey-Berino J, Pintauro S, Buzzell P, DiGiulio M, Casey Gold B, Moldovan C, et al. Does using the Internet facilitate the maintenance of weight loss? Int J Obes Relat Metab Disord 2002 Sep;26(9):1254-1260 [FREE Full text] [Medline: 12187404] [doi: 10.1038/sj.ijo.0802051]

24. Oenema A, Brug J, Lechner L. Web-based tailored nutrition education: results of a randomized controlled trial. Health Educ Res 2001 Dec;16(6):647-660 [FREE Full text] [Medline: 11780705] [doi: 10.1093/her/16.6.647]
25. Marshall AL, Leslie ER, Bauman AE, Marcus BH, Owen N. Print versus website physical activity programs: a randomized trial. Am J Prev Med 2003 Aug;25(2):88-94. [Medline: 12880874] [doi: 10.1016/S0749-3797(03)00111-9]
26. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [Medline: 16908795] [doi: 10.1001/archinte.166.15.1620]
27. 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]
28. 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]
29. Womble LG, Wadden TA, McGuckin BG, Sargent SL, Rothman RA, Krauthamer-Ewing ES. A randomized controlled trial of a commercial internet weight loss program. Obes Res 2004 Jun;12(6):1011-1018 [FREE Full text] [Medline: 15229342]

submitted 08.12.06; peer-reviewed by K Rothert, J Harvey-Berino; comments to author 16.05.07; revised version received 22.05.07; accepted 22.05.07; published 19.06.07.

Please cite as:
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
URL: http://www.jmir.org/2007/2/e17/
doi:10.2196/jmir.9.2.e17
PMID:

Original Paper

Internet-Based Interactive Health Intervention for the Promotion of Sensible Drinking: Patterns of Use and Potential Impact on Members of the General Public

Stuart Linke[1], MSc, AFBPsS; Elizabeth Murray[2], FRCGP, PhD; Ceri Butler[2], MSc; Paul Wallace[2], FRCGP, FFPHM

[1]Islington Mental Health Psychology Services, Camden and Islington Mental Health and Social Care Trust, London, UK

[2]Primary Care and Population Sciences, University College London, London, UK

Corresponding Author:
Stuart Linke, MSc, AFBPsS
Head of Islington Mental Health Psychology Services
Camden and Islington Mental Health and Social Care Trust
5th Floor, Hill House, 17 Highgate Hill
London N19 5NA
United Kingdom
Phone: +44 020 7685 4700
Fax: +44 020 7685 4786
Email: stuart.linke@candi.nhs.uk

Abstract

Background: Heavy drinking is responsible for major health and social problems. Brief interventions have been shown to be effective, but there have been difficulties in reaching those who might benefit from them. Pilot studies have indicated that a Web-based intervention is likely to be acceptable to heavy drinkers and may produce some health benefits. However, there are few data on how many people might use such a program, the patterns of use, and potential benefits.

Objectives: The aim was to examine the demographic characteristics of users of a free, Web-based, 6-week intervention for heavy drinkers and to describe the methods by which users identified the site, the pattern of site use and attrition, the characteristics associated with completing the program, and the self-reported impact on alcohol-related outcomes.

Methods: Cohort study. Visitors to the Web site were offered screening with the Fast Alcohol Screening Test, and those scoring above the cutoff for risky drinking were invited to register with the program. Demographic information was collected routinely at registration, and questionnaires were completed at the end of weeks 1 and 6. The outcome measures assessed dependency (Short Alcohol Dependency Data Questionnaire), harms (modified Alcohol Problems Questionnaire), and mental health (Clinical Outcomes in Routine Evaluation–Outcome Measure).

Results: The records of 10000 users were analyzed. The mean age was 37.4 years, 51.1% were female, 37.5% were single, and 42.4% lived with children. The majority were White British, lived in the United Kingdom, and reported occupations from the higher socioeconomic strata. Over 70% connected to the Down Your Drink (Down Your Drink) site from another Internet-based resource, whereas only 5.8% heard about the site from a health or other professional. Much of the Web site use (40%) was outside normal working hours. Attrition from the program was high, with only 16.5% of registrants completing the whole 6 weeks. For those who completed the program, and the final outcome measures, measures of dependency, alcohol-related problems, and mental health symptoms were all reduced at week 6.

Conclusions: The Web-based intervention was highly used, and those who stayed with the program showed significant reductions in self-reported indicators of dependency, alcohol-related problems, and mental health symptoms; however, this association cannot be assumed to be causal. Programs of this type may have the potential to reach large numbers of heavy drinkers who might not otherwise seek help. There are significant methodological challenges and further research is needed to fully evaluate such interventions.

(J Med Internet Res 2007;9(2):e10) doi:10.2196/jmir.9.2.e10

KEYWORDS

Alcohol; drinking; health promotion; Internet

Introduction

Excess alcohol consumption, and the harm caused by it, is a major public health concern throughout the developed world [1-3]. Regular heavy alcohol consumption and binge drinking are associated with physical problems, mental health problems, antisocial behavior, violence, accidents, suicide, injuries, road traffic accidents, unsafe sexual behavior, underperformance at school, and crime. In the United Kingdom, alcohol misuse accounts for over 30000 hospital admissions for alcohol dependence, up to 70% of all admissions to accident and emergency departments at peak times, and up to 22000 premature deaths. The total cost of alcohol misuse to the health service was calculated to be £1.7 billion per annum, which though substantial, is much less than the total annual cost of alcohol-related crime (£7.4 billion) and lost productivity (£6.4 billion) [3]. Similar costs are identified in the United States, with the overall economic cost of alcohol abuse estimated at $184.6 billion, most of which was attributed to lost productivity [1]. Alcohol dependence is not confined to adulthood: in 2000, nearly 14% of 16- to 19-year-olds in Great Britain were found to experience dependence on alcohol [3].

Brief interventions seek to change views of the personal acceptability of excessive drinking and to encourage self-directed behavior change. They can be delivered by practitioners or as self-help materials. There is a substantial body of evidence demonstrating that brief interventions for individuals at risk can have significant impact on reducing alcohol consumption and, in some cases, alcohol-related harm when delivered both in primary and secondary health care settings [4-8]. However, their impact on public health has been severely limited, due partly to the reluctance of at-risk individuals to seek help [9] and partly to a lack of health care resources and the unwillingness of health care professionals to undertake these interventions [10]. General practitioners, for example, who are in a key position to deliver such interventions, rarely do so because of both lack of skills and fear of potential adverse effects on relationships with their patients [10].

Until recently, self-help materials were almost exclusively paper-based. However, the Internet has triggered a growth in more interactive self-help materials. It is thought that this interactivity is likely to enhance the potential for behavior change [11]. In public health terms, the Internet has the considerable advantage that once the Web site has been developed, the marginal cost of delivering the intervention to unlimited numbers of people is minimal, in marked contrast to face-to-face interventions. Approximately 60% of individuals in the United Kingdom report using the Internet regularly [12], and a recent telephone survey in Canada showed that current drinkers are more likely to have access to the Internet than abstainers [13]. Heavy drinkers have been shown to benefit from a PC Windows-based behavioral treatment program [14,15]. Younger people, who are most at risk of binge drinking [3], use new information and communication technologies such as the Internet and mobile phones in preference to more traditional sources of health information or health promotion [16]. Pilot studies have shown that people with drinking problems are willing to use screening tools on the Internet [17,18]. One study asked Web site visitors to complete a questionnaire about their alcohol use and provided a brief intervention in the form of feedback and advice [19]. A recent review identified a number of Internet-based screening and assessment tools (but few treatment or intervention-based applications), which were mostly directed at college students and usually required users to attend special sessions in an office [20-22].

One exception is the Alcohol Help Centre, which provided online personal feedback to users of an online eHealth service [23]. A modified version of this Internet-based tool was piloted in Finland. In a relatively small cohort study (n = 343), at 3-month follow-up, users had reduced their drinking compared to baseline [24].

Down Your Drink is a well-established, comprehensive, freely available, interactive Web-based treatment program for people with alcohol problems. An initial pilot study demonstrated the feasibility of the approach [25], and this present study seeks to extend what is known about the “natural history” of the site over a longer period and to describe the users and their outcomes. All advertising and promotion of the site ceased after the pilot study, but the Web-based program continued to be hosted on the Web site of Alcohol Concern, the United Kingdom’s premier charity addressing issues around alcohol.

The aims of the present study were to describe the patterns of use and self-reported effectiveness among users of Down Your Drink. The study set out to describe the demographic characteristics of users, the methods employed to identify and access Down Your Drink, the patterns of use, the demographic and clinical characteristics associated with completing the 6-week program, and self-reported changes in alcohol-related outcomes associated with use of Down Your Drink.

Methods

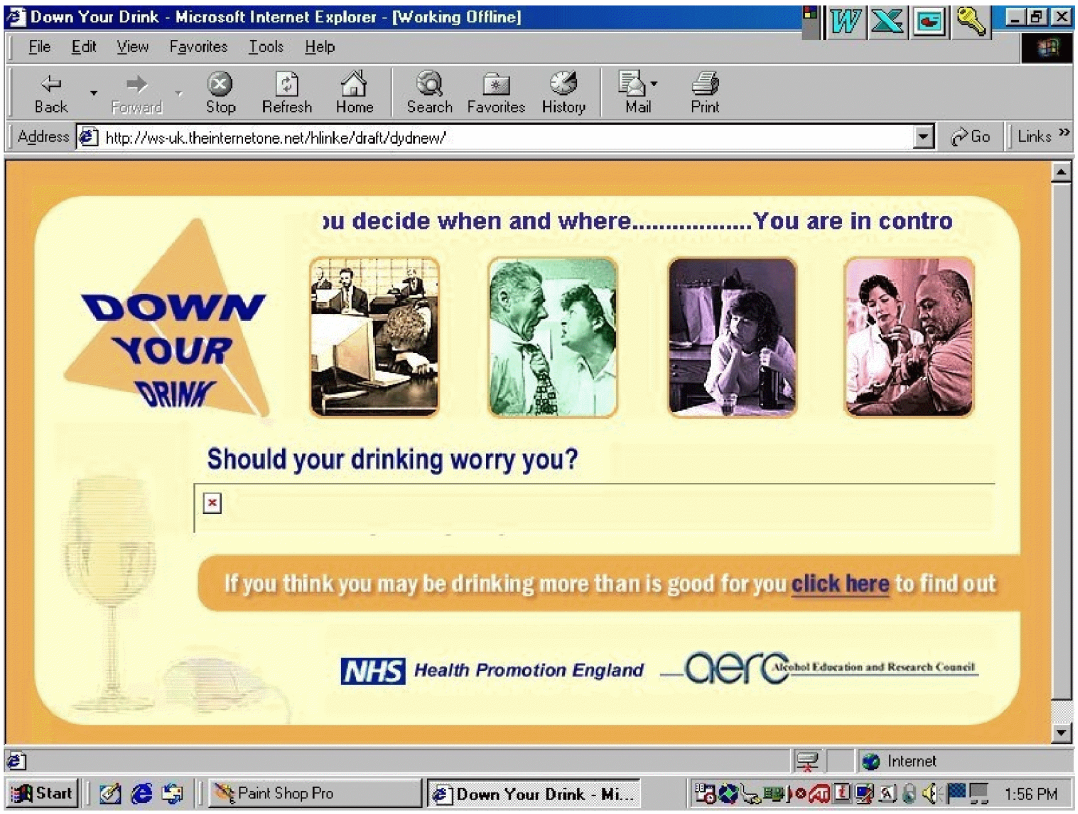
This was a pragmatic cohort study of the first 10000 people who registered to use the site, after the end of the pilot phase (ie, after September 2003).

The study was approved by the Camden and Islington Local Research Ethics Committee.

Down Your Drink was developed with support from the Alcohol Education and Research Council as a Web-based interactive program of brief interventions to reduce alcohol consumption in heavy drinkers and is hosted on a single dedicated Web site [25].

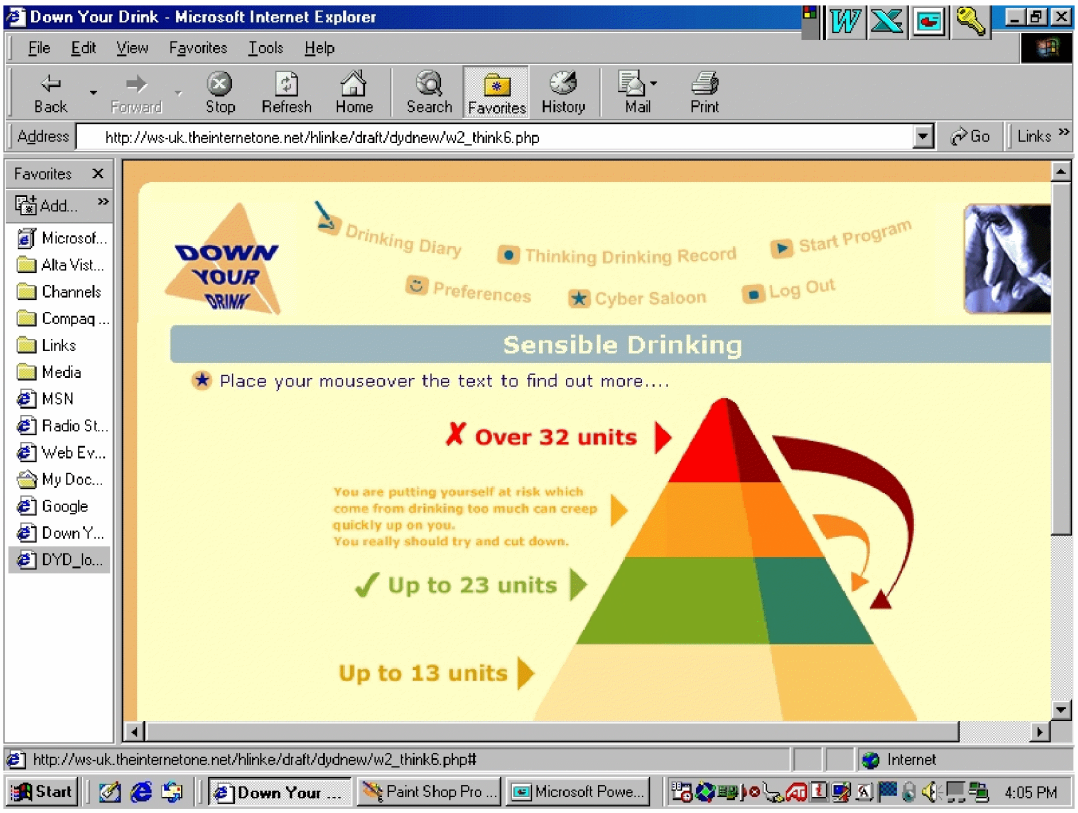
The animated home page invited visitors to assess their level of drinking by taking the Fast Alcohol Screening Test (FAST) [26]. Feedback was presented numerically and visually. The home page (Figure 1) also contained links to frequently asked questions (FAQs) about heavy drinking, information about the 6-week program, its authors, the research study, and the organizations sponsoring the site (The National Health Service, Alcohol Concern, Alcohol Education and Research Council, and the Royal Free and University College London Medical School).

Figure 1. Down Your Drink home page



The Web site content was structured as a set of six consecutive intervention modules designed to be accessed by the registrants at weekly intervals. Once a given module had been completed, access to the subsequent module was barred for 7 days. The 6-week program was based on the stages of change model [27] and contained components common in brief interventions, including motivational enhancement [28], cognitive behavioral therapy [29], and relapse prevention [30]. Experience with the original print version of the program suggested that structuring the material over 6 weeks allowed users sufficient time to complete the various items and plan their behavior change. There were also components unique to a computer-based system, designed to exploit the interactivity and flexibility of the Internet (Figure 2 and Multimedia Appendix 1). These included an automated drinking diary and consumption calculator, online quizzes, interactive behavioral analysis of drinking situations (the “thinking drinking” log), blood alcohol concentration calculator, and intelligent email that sent reminders and controlled-drinking tips to an individual email address or as a Short Message Service (SMS) text message to a cell phone. Associated with the Down Your Drink program was a nonmoderated listserve for Down Your Drink users hosted by Yahoo groups where users of the program could exchange personal messages about their experience with the program and obtain peer support for their efforts in reducing their drinking.

Figure 2. Sample Down Your Drink page



Recruitment

An off-line advertising campaign was run in September 2001 to coincide with the launch of the Web site and the early part of the pilot study. All off-line advertising ceased at the end of 2001 (Multimedia Appendix 2). The site was also registered with the Yahoo search engine and was listed in the information pages of a number of UK-based health-related periodicals. By the time this study started (September 2003), there had been no off-line advertising for this Web site for over 18 months. All visitors to the Down Your Drink site were invited to complete an initial assessment using the FAST [26]. The maximum score for the FAST is 16, and the cutoff point for risky drinking is 3. Those who scored 3 or above were advised that they were at risk from their drinking behavior and were directed to the free 6-week program. Those with lower FAST scores were also able to use the program. Visitors who wished to participate in the 6-week program were required to complete an online consent form (Multimedia Appendix 3) and provide registration data prior to accessing the site. Following this, they could enter the first week of the program.

Baseline and Outcome Measures

In order to participate, users were invited to read the policy on confidentiality, complete the consent procedure, and then choose a username and password. Those who agreed to register were required to submit information about their age, gender, marital status, family composition, ethnicity, occupation, country of residence, and how they had learned about the site.

The primary outcome measure was the 14-item Short Alcohol Dependency Questionnaire (SADD), which measures dependency on alcohol [31]. This was selected in preference to a measure of alcohol consumption as, at that time, there was no direct measure of alcohol use validated for use on the Internet. Secondary outcome measures included an abbreviated (35-item) version of the Alcohol Problems Questionnaire (APQ), which assesses harm associated with heavy drinking [32], and the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM), a 34-item measure of mental health symptoms [33]. For all these questionnaires, higher scores indicate greater levels of harm. All measures were completed online at the end of weeks 1 and 6. Respondents were able to review and check their responses by the using the back button on their browser. Respondents were not required to complete all the items in each questionnaire, thus allowing the possibility of differential response rates on each questionnaire.

All the data for the study were collected automatically on a live database located on a secure dedicated Web site. Access to this database was by password only and was restricted to members of the research team. This database was initially launched simultaneously with the Down Your Drink Web site in October 2001, but was withdrawn at the end of the pilot phase for re-development. The upgraded database was subsequently

re-launched in September 2003, and the data presented here are from the first 10000 users after the launch of the revised database.

Usage data were collected automatically on the Down Your Drink Web server and analyzed by a Web server log file analysis program (Webalizer, version 2.01). These data were stored separately from the Down Your Drink registrant database. The usage data reported here are from a subset of the total participants in the study collected between January and December 2004.

Analysis

Data for the first 10000 users were extracted from the live database on March 1, 2006, and transferred into SPSS, version 12.0.1 for Windows via Microsoft Access, where all data related to individual Web site users were linked using the unique identifier created by the Down Your Drink site.

These data were then subjected to frequency analysis and paired *t* tests were performed to compare the mean scores of Web site users at the beginning and end of the 6-week program. In addition, the demographic variables of age, gender, marital status (with or without partner), and family status (with or without children) were tested using a chi-square test to determine if a statistically significant relationship existed between particular demographic characteristics and the likelihood of completing the online program.

Results

It took just over 27 months, from the launch date on September 24, 2003, to January 3, 2006, to complete recruitment of 10000 users.

Demographic Characteristics of Down Your Drink Users

The self-reported demographic characteristics of the sample are shown in Table 1. It can be seen that approximately equal numbers of men and women used the program; a little over a third were single, and over 40% lived with children. Most users (81.9%) were White British, and over a quarter reported managerial and professional occupations.

The great majority (83.9%) lived in the United Kingdom, and 9.3% reported living in other English-speaking countries (United States, Canada, Australia, and New Zealand). Over 100 countries of residence were given by the remaining 6.7% of users.

Table 1. Demographic characteristics of Down Your Drink users

Demographic Characteristic		
	Mean	**SD**
Age	37.44	9.84
Gender	**No.**	**%**
Male	4891	48.9
Female	5109	51.1
Marital Status		
Single	3754	37.5
Married or living with partner	6246	62.5
Living With Children		
Yes	4244	42.4
No	5756	57.6
Occupation		
Managerial/professional	2579	25.8
Self-employed	862	8.6
Administrative/secretarial	854	8.5
Information technology	770	7.7
Academic	461	4.6
Housewife/househusband	431	4.3
Unemployed	353	3.5
All other	3690	36.9
Ethnicity		
White British	8185	81.9
White other	911	9.1
White Irish	527	5.3
Asian	135	1.4
Mixed	93	0.9
Black	69	0.7
Other	81	0.8
Country of Origin		
United Kingdom	8385	83.9
United States	554	5.5
Ireland	172	1.7
Australia	111	1.1
Canada	104	1.0
Other	674	6.7

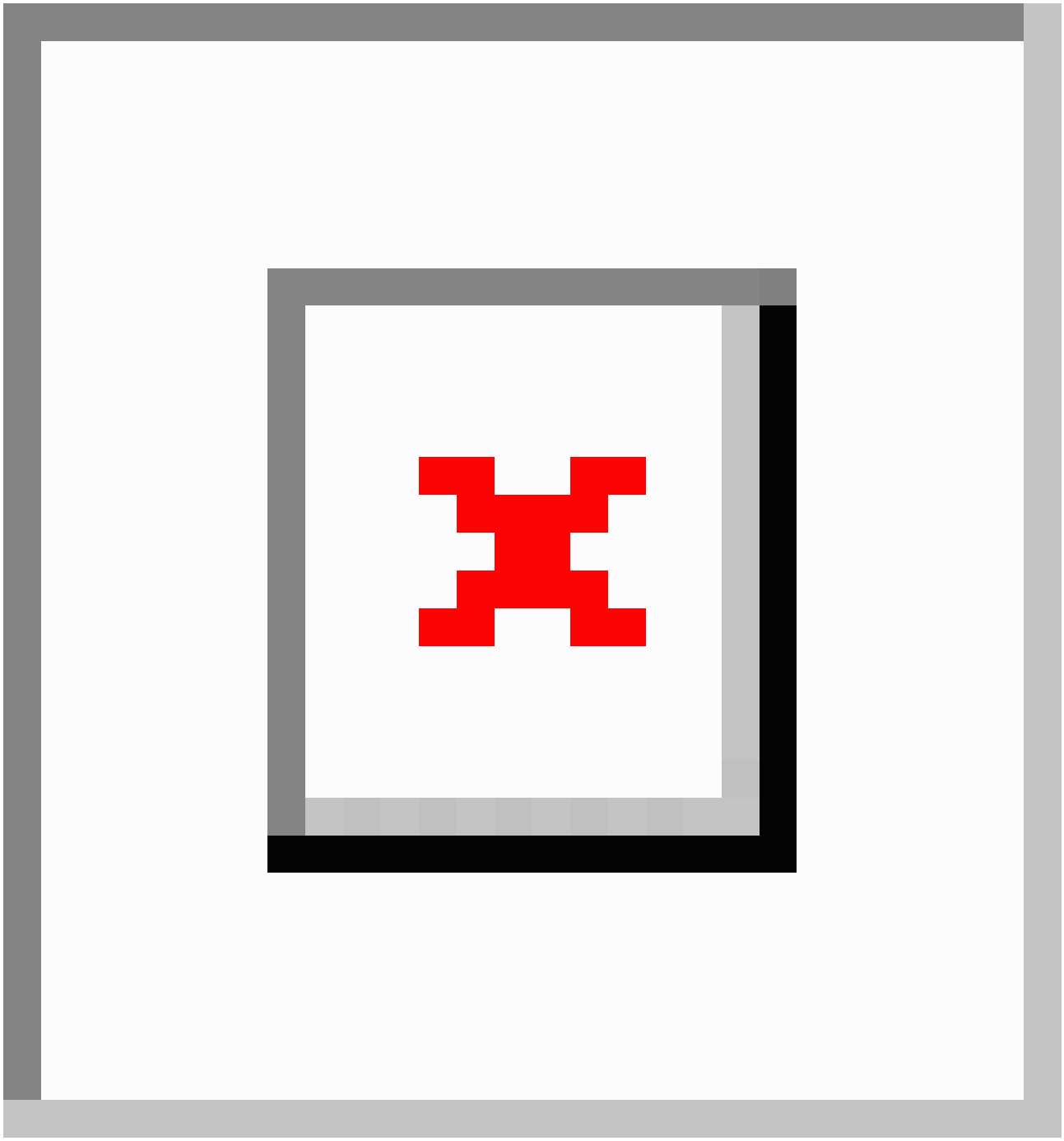
How Users Found Down Your Drink

Nearly three quarters of registrants (n = 7167, 71.7%) visited the Down Your Drink site from another Internet-based resource. Most of these connected via a link from another Web site (n = 4156, 41.6%) or from a search engine (n = 2900, 29.0%), whereas a smaller group responded to a banner advert on another site (n = 111, 1.0%). Relatively few registrants had been directed to Down Your Drink from the health service (n = 583, 5.8%).

When Users Accessed Down Your Drink

The daily pattern of Down Your Drink use between January and December 2004 (represented by number of "hits" per hour) is shown in Figure 3. About 61% of the hits occurred between 9:00 am and 5:59 pm (Greenwich mean time [GMT]), with about 39% occurring between 6:00 pm and 8:59 am.

Figure 3. Mean hourly usage of Down Your Drink (Jan to Dec 2004)



Use of Down Your Drink

There was a high attrition rate from the program with only 1654 (16.5%) of the original 10000 registrants completing the 6-week program (Table 2). The greatest attrition occurred between the first and the second weeks, with much lower rates occurring thereafter.

Table 2. Number of users completing each week of the program. Percentages are the proportion of those registered, for example, 16.5% of all users that registered with Down Your Drink had completed the 6-week program by the time of the data extraction (March 1, 2006).

	All Users		Male Users		Female Users	
	No.	%	No.	%	No.	%
Registered	10000	100.0	4891	100	5109	100
Completed week 1	8933	89.3	4302	88.0	4631	90.6
Completed week 2	4020	40.2	1916	39.2	2104	41.2
Completed week 3	3006	30.1	1403	28.7	1603	31.4
Completed week 4	2411	24.1	1128	23.1	1283	25.1
Completed week 5	1928	19.3	887	18.1	1041	20.4
Completed week 6	1654	16.5	770	15.7	884	17.3

Outcomes

In order to determine whether there were demographic or clinical characteristics that were associated with completion of the 6-week program, we compared the characteristics of those who completed the sixth week with those who had only completed the first week. We chose first week completers as the reference, as we did not have data on the outcome measures from those who had dropped out before completing the first week. We defined anyone who had done some part of week 6 as a completer, whether or not they filled in the questionnaires at the end of week 6. It can be seen that female users, users who were married or living with a partner, and users without children were more likely to complete the program than men, single users, or users with children (Table 3). The baseline responses of those users who completed the program indicated that they were less at risk of alcohol dependency and of harm from alcohol use at the time of entry into the program than those who subsequently dropped out (Table 3), although they were still at considerable risk of harm as judged by their baseline scores.

Table 3. Comparison of baseline demographic and clinical characteristics of those who completed the six week program (completers) with those who only completed the first week of the program (starters)

Characteristic	Starters (Completed Week 1) (n = 8933)	Completers (Completed Week 6) (n = 1654)	*P* value
Demographic Characteristic	**No. (%)**	**No. (%)**	
Gender			0.04
Male	4302 (48)	770 (47)	
Female	4631 (52)	884 (53)	
Marital Status			< .001
Married or living with partner	5646 (63)	1125 (68)	
Single	3287 (37)	529 (32)	
Family Status			.002
With children	3808 (43)	646 (39)	
Without children	5125 (57)	1008 (61)	
Age	**Mean (SD)**	**Mean (SD)**	
Years	37.6 (9.8)	38.9 (9.6)	
Clinical Characteristic	**Mean (SD)**	**Mean (SD)**	
SADD*	12.31 (6.09)	11.52 (5.24)	< .001
Abbreviated APQ†	7.38 (5.01)	6.83 (4.66)	< .001
Core functioning	1.38 (0.82)	1.35 (0.74)	.17
Core problem	1.59 (0.92)	1.59 (0.81)	.93
Core well-being	1.62 (1.00)	1.57 (0.92)	.06
Core risk	0.41 (0.60)	0.34 (0.50)	< .001

*Short Alcohol Dependency Data Questionnaire

†Alcohol Problems Questionnaire

Of the program completers, 57% also fully completed the outcome questionnaires at the end of the 6-week program. The mean scores for the SADD and modified APQ were significantly lower for these people at week 6 than week 1, indicating that alcohol dependency and alcohol-related harm were significantly reduced at the end of the program (Table 4). Mental health symptoms were also significantly improved. At week 1, the mean item scores for the CORE-OM were at or above the clinical thresholds on all domains for both women and men, with the single exception of the risk score for men. At week 6, the mean item scores were significantly reduced across all scales, suggesting a reduction in mental health symptoms, a reduced risk of harm to self and others, and an improvement in subjective well-being and daily functioning (Table 4).

Table 4. Change in clinical outcomes between week 1 and week 6 in users who completed the 6-week program

Clinical Outcome Measure	Week 1 (Baseline) Score Mean (SD)	Week 6 (Final) Score Mean (SD)	*P* value
SADD (total score)			
Men (n = 421)	11.51 (5.17)	7.65 (4.51)	< .001
Women (n = 520)	11.58 (5.35)	7.64 (5.04)	< .001
APQ (total score)			
Men (n = 421)	7.18 (4.74)	3.43 (3.90)	< .001
Women (n = 520)	6.61 (4.35)	3.05 (3.66)	< .001
CORE-OM (item scores)			
Men (n= 421)			
Functioning	1.32 (0.73)	0.87 (0.69)	< .001
Problem	1.51 (0.81)	0.94 (0.74)	< .001
Well-being	1.41 (0.90)	0.88 (0.84)	< .001
Risk	0.31 (0.46)	0.14 (0.33)	< .001
Women (n = 520)			
Functioning	1.37 (0.75)	0.85 (0.72)	< .001
Problem	1.66 (0.80)	0.99 (0.80)	< .001
Well-being	1.70 (0.91)	1.00 (0.89)	< .001
Risk	0.38 (0.55)	0.18 (0.44)	< .001

Discussion

Main Findings

This large pragmatic cohort study of users of a freely available online program for nondependent drinkers at risk of harm from alcohol use suggests that a small but significant (16.5% or 1 in 6) proportion of users will complete the 6-week program. Those that completed the program and provided outcome measures reported clinically significant benefit, with reduction in mean dependency and harm from alcohol, and improved mental health. Women, users with a partner, and users without children were more likely to complete the program.

Relationship to Previous Literature

Our findings add to the growing body of literature suggesting that behavior change can be achieved through online interventions [34-36]. This suggestion provides some empirical support for the move toward providing online public health interventions [37]. In common with many other online interventions [38], we had a very high rate of attrition. As Eysenbach has pointed out, this may be integral to the nature of online interventions and does not undermine their potential usefulness, as the marginal cost per user is so low. We also found that the online intervention was frequently accessed outside normal working hours [39], although our findings on this point must be treated with caution, as a proportion of our users came from outside the United Kingdom, and the time of access was recorded using GMT.

This online service, in common with other studies [24,40], was used by as many women as men. This is an important difference from traditional alcohol treatment services, where women tend to be underrepresented and report that the services they receive do not meet their needs [41].

Methodological Issues

This was a pragmatic cohort study aimed at exploring the usage patterns of a freely available online intervention. All data were self-reported, and we have no objective confirmation of any of the data reported here. Some users may have provided false or inaccurate data. However, there is no particular reason why registrants should have systematically lied at registration, and, given the large sample size (10000 registrants), the demographic data probably provide a reasonable description of the characteristics of users. Users who completed the 6-week course showed considerable motivation and commitment. The outcome measures were an integral part of the intervention, and users completed them entirely for their own benefit. It is difficult to know how truthful they were in completing these questionnaires, but there is some evidence to support the suggestion that responses to online questionnaires are comparable to traditional paper-based versions [42,43].

As this was an uncontrolled study, the data can only suggest an association between use of the program and an improvement in health outcomes. This study cannot determine whether this association was causal.

Implications

If confirmed, these data suggest that an online intervention aimed at nondependent heavy drinkers can make a useful addition to the public health armamentarium. Although only 16% of those who registered completed the course, in public health terms this still represents a significant number of people who could benefit. The advantages of such an intervention are that the costs are unaffected by the number of users and that the intervention can be used at home, at any time of day or night, unlike more traditional services. Future work should explore whether adapting the program, including increasing the options for flexible use, removing the requirement to work through the elements sequentially, and introducing more personalized feedback, can improve "stickiness" or adherence to the program, and whether the association between use of the program and improved outcomes is causal. We are currently undertaking a randomized controlled trial to explore both these questions.

Acknowledgments

The development of Down Your Drink was supported by grants from the Alcohol Education and Research Council and the Department of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of the Down Your Drink program

[PDF file (Adobe Acrobat), 2.7 MB - jmir_v9i2e10_app1.pdf]

Multimedia Appendix 2

Recruitment leaflet sent to health care professionals

[PDF file (Adobe Acrobat), 96 KB - jmir_v9i2e10_app2.pdf]

Multimedia Appendix 3

Participant information sheet

[PDF file (Adobe Acrobat), 92 KB - jmir_v9i2e10_app3.pdf]

References

1. ; Department of Health and Human Services. 10th Special Report to the US Congress on Alcohol and Health. Washington, DC: Department of Health and Human Services; 2000. URL: http://pubs.niaaa.nih.gov/publications/10report/intro.pdf [WebCite Cache ID 5OatcGn2R]
2. ; The Council of the European Union. Council conclusions of 5 June 2001 on a community strategy to reduce alcohol-related harm. Off J European Communities 2001;175:1.
3. ; Cabinet Office Strategy Unit. Alcohol Harm Reduction Strategy for England. London, UK: Prime Minister's Strategy Unit; 2004. URL: http://image.guardian.co.uk/sys-files/Society/documents/2004/03/15/alcoholstrategy.pdf [WebCite Cache ID 5ObIdlWdN]
4. Ballesteros J, Duffy JC, Querejeta I, Ariño J, González-Pinto A. Efficacy of brief interventions for hazardous drinkers in primary care: systematic review and meta-analyses. Alcohol Clin Exp Res 2004 Apr;28(4):608-618. [Medline: 15100612] [doi: 10.1097/01.ALC.0000122106.84718.67]
5. Beich A, Thorsen T, Rollnick S. Screening in brief intervention trials targeting excessive drinkers in general practice: systematic review and meta-analysis. BMJ 2003 Sep 6;327(7414):536-542 [FREE Full text] [Medline: 12958114] [doi: 10.1136/bmj.327.7414.536]
6. Bien TH, Miller WR, Tonigan JS. Brief interventions for alcohol problems: a review. Addiction 1993 Mar;88(3):315-335. [Medline: 12958114] [doi: 10.1111/j.1360-0443.1993.tb00820.x]
7. Heather N. Effectiveness of brief interventions proved beyond reasonable doubt. Addiction 2002 Mar;97(3):293-294. [Medline: 11964102] [doi: 10.1046/j.1360-0443.2002.0105a.x]
8. Moyer A, Finney JW, Swearingen CE, Vergun P. Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. Addiction 2002 Mar;97(3):279-292. [Medline: 11964101] [doi: 10.1046/j.1360-0443.2002.00018.x]
9. Heather N. The public health and brief interventions for excessive alcohol consumption: the British experience. Addict Behav 1996 Nov;21(6):857-868. [Medline: 8904949] [doi: 10.1016/0306-4603(96)00040-8]

10. Beich A, Gannik D, Malterud K. Screening and brief intervention for excessive alcohol use: qualitative interview study of the experiences of general practitioners. BMJ 2002 Oct 19;325(7369):870 [FREE Full text] [Medline: 12386040] [doi: 10.1136/bmj.325.7369.870]
11. Ritterband L, Gonder-frederick LA, Cox DJ, Clifton AD, West RW, Borowitz S. Internet interventions: in review, in use, and into the future. Prof Psychol Res Pract 2003;34(5):527-534. [doi: 10.1037/0735-7028.34.5.527]
12. Dutton WH, di Gennaro C, Hargrave AM. The Oxford Internet Survey (OxIS) Report 2005: The Internet in Britain. Oxford, UK: Oxford Internet Institute; 2005.
13. Cunningham JA, Selby PL, Kypri K, Humphreys KN. Access to the Internet among drinkers, smokers and illicit drug users: is it a barrier to the provision of interventions on the World Wide Web? Med Inform Internet Med 2006 Mar;31(1):53-58. [Medline: 16754367] [doi: 10.1080/14639230600562816]
14. Hester RK, Delaney HD. Behavioral Self-Control Program for Windows: results of a controlled clinical trial. J Consult Clin Psychol 1997 Aug;65(4):686-693. [Medline: 9256570] [doi: 10.1037/0022-006X.65.4.686]
15. Squires DD, Hester RK. Using technical innovations in clinical practice: the Drinker's Check-Up software program. J Clin Psychol 2004 Feb;60(2):159-169. [Medline: 14724923] [doi: 10.1002/jclp.10242]
16. Rideout V. Generation Rx.com: How Young People Use the Internet for Health Information. Menlo Park, CA: Kaiser Family Foundation; 2001. URL: www.kff.org/entmedia/upload/Toplines.pdf [WebCite Cache ID 5OattXlVw]
17. Cloud RN, Peacock PL. Internet screening and interventions for problem drinking: results from the www.carebetter.com pilot study. Alcohol Treat Q 2001;19(2):23-44. [doi: 10.1300/J020v19n02_02]
18. Cunningham JA, Humphreys K, Koski-Jännes A. Providing personalized assessment feedback for problem drinking on the Internet: a pilot project. J Stud Alcohol 2000 Nov;61(6):794-798. [Medline: 11188484]
19. Saitz R, Helmuth ED, Aromaa SE, Guard A, Belanger M, Rosenbloom DL. Web-based screening and brief intervention for the spectrum of alcohol problems. Prev Med 2004 Nov;39(5):969-975. [Medline: 15475031] [doi: 10.1016/j.ypmed.2004.04.011]
20. Hester RK, Miller JH. Computer-based tools for diagnosis and treatment of alcohol problems. Alcohol Res Health 2006;29(1):36-40 [FREE Full text] [Medline: 16767852]
21. Hester RK, Squires DD, Delaney HD. The Drinker's Check-up: 12-month outcomes of a controlled clinical trial of a stand-alone software program for problem drinkers. J Subst Abuse Treat 2005 Mar;28(2):159-169. [Medline: 15780546] [doi: 10.1016/j.jsat.2004.12.002]
22. Kypri K, Saunders JB, Williams SM, Mcgee RO, Langley JD, Cashell-Smith ML, et al. Web-based screening and brief intervention for hazardous drinking: a double-blind randomized controlled trial. Addiction 2004 Nov;99(11):1410-1417. [Medline: 15500594] [doi: 10.1111/j.1360-0443.2004.00847.x]
23. Cunningham JA, Humphreys K, Kypri K, Van Mierlo T. Formative evaluation and three-month follow-up of an online personalized assessment feedback intervention for problem drinkers. J Med Internet Res 2006 Apr;8(2):e5 [FREE Full text] [Medline: 16867968] [doi: 10.2196/jmir.8.2.e5]
24. Koski-Jännes A, Cunningham JA, Tolonen K, Bothas H. Internet-based self-assessment of drinking--3-month follow-up data. Addict Behav 2007 Mar;32(3):533-542. [Medline: 16822619] [doi: 10.1016/j.addbeh.2006.05.020]
25. Linke S, Brown A, Wallace P. Down your drink: a web-based intervention for people with excessive alcohol consumption. Alcohol Alcohol 2004;39(1):29-32 [FREE Full text] [Medline: 14691071] [doi: 10.1093/alcalc/agh004]
26. Hodgson RJ, John B, Abbasi T, Hodgson RC, Waller S, Thom B, et al. Fast screening for alcohol misuse. Addict Behav 2003 Oct;28(8):1453-1463 [FREE Full text] [Medline: 14512068] [doi: 10.1016/S0306-4603(02)00246-0]
27. Prochaska JO, Diclemente CC, Velicer WF, Ginpil S, Norcross JC. Predicting change in smoking status for self-changers. Addict Behav 1985;10(4):395-406. [Medline: 6863699] [doi: 10.1016/0306-4603(85)90036-X]
28. Dunn C, Deroo L, Rivara FP. The use of brief interventions adapted from motivational interviewing across behavioral domains: a systematic review. Addiction 2001 Dec;96(12):1725-1742. [Medline: 11784466] [doi: 10.1080/09652140120089481]
29. Monti P, Kadden R, Rohsenow D, Cooney N, Abrams D. Treating Alcohol Dependence: A Coping Skills Training Guide. New York: Guildford Press; 2002.
30. Marlatt GA, Gordon JR. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors. New York: Guildford Press; 1985.
31. Davidson R, Raistrick D. The validity of the Short Alcohol Dependence Data (SADD) Questionnaire: a short self-report questionnaire for the assessment of alcohol dependence. Br J Addict 1986 Apr;81(2):217-222. [Medline: 3458489] [doi: 10.1111/j.1360-0443.1986.tb00319.x]
32. Williams BT, Drummond DC. The Alcohol Problems Questionnaire: reliability and validity. Drug Alcohol Depend 1994 Jun;35(3):239-243. [Medline: 7956754] [doi: 10.1016/0376-8716(94)90080-9]
33. Evans C, Connell J, Barkham M, Margison F, Mcgrath G, Mellor-Clark J, et al. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. Br J Psychiatry 2002 Jan;180(1):51-60 [FREE Full text] [Medline: 11772852] [doi: 10.1192/bjp.180.1.51]
34. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [Medline: 16235356] [doi: 10.1002/14651858.CD004274.pub4]

35. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, Mcghee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [Medline: 15631964] [doi: 10.2196/jmir.6.4.e40]
36. Nguyen HQ, Carrieri-Kohlman V, Rankin SH, Slaughter R, Stulbarg MS. Internet-based patient education and support interventions: a review of evaluation studies and directions for future research. Comput Biol Med 2004 Mar;34(2):95-112. [Medline: 14972630] [doi: 10.1016/S0010-4825(03)00046-5]
37. Marshall AL, Owen N, Bauman AE. Mediated approaches for influencing physical activity: update of the evidence on mass media, print, telephone and website delivery of interventions. J Sci Med Sport 2004 Apr;7(1 Suppl):74-80. [Medline: 15214605] [doi: 10.1016/S1440-2440(04)80281-0]
38. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [Medline: 15829473] [doi: 10.2196/jmir.7.1.e11]
39. Brennan PF, Moore SM, Smyth KA. Alzheimer's disease caregivers' uses of a computer network. West J Nurs Res 1992 Oct;14(5):662-673. [Medline: 1529609] [doi: 10.1177/019394599201400508]
40. Humphreys K, Klaw E. Can targeting nondependent problem drinkers and providing internet-based services expand access to assistance for alcohol problems? A study of the moderation management self-help/mutual aid organization. J Stud Alcohol 2001 Jul;62(4):528-532. [Medline: 11513231]
41. Copeland J, Hall W. A comparison of predictors of treatment drop-out of women seeking drug and alcohol treatment in a specialist women's and two traditional mixed-sex treatment services. Br J Addict 1992 Jun;87(6):883-890. [Medline: 1326359] [doi: 10.1111/j.1360-0443.1992.tb01983.x]
42. 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]
43. Miller ET, Neal DJ, Roberts LJ, Baer JS, Cressler SO, Metrik J, et al. Test-retest reliability of alcohol measures: is there a difference between internet-based assessment and traditional methods? Psychol Addict Behav 2002 Mar;16(1):56-63. [Medline: 11934087] [doi: 10.1037/0893-164X.16.1.56]

Abbreviations

APQ: Alcohol Problems Questionnaire
CORE-OM: Clinical Outcomes in Routine Evaluation-Outcome Measure
FAST: Fast Alcohol Screening Test
GMT: Greenwich mean time
SADD: Short Alcohol Dependency Questionnaire

submitted 10.07.06; peer-reviewed by J Cunningham; comments to author 20.12.06; revised version received 21.02.07; accepted 30.03.07; published 08.05.07.

Please cite as:
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
URL: http://www.jmir.org/2007/2/e10/
doi:10.2196/jmir.9.2.e10
PMID:17513281

Original Paper

Language Preferences on Websites and in Google Searches for Human Health and Food Information

Punam Mony Singh[1], MD, FRCPC; Carly A Wight[2]; Olcan Sercinoglu[3], BSc; David C Wilson[4], BEng; Artem Boytsov[3], BSc; Manish N Raizada[5], PhD

[1]Respirology Program, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[2]Croplink Global Food Security Initiative, Department of Plant Agriculture, University of Guelph, Guelph, ON, Canada

[3]Google Inc, Mountain View, CA, USA

[4]Department of Psychology, University of Guelph, Guelph, ON, Canada

[5]Croplink Global Food Security Initiative, Department of Plant Agriculture, University of Guelph, Guelph, ON, Canada

Corresponding Author:
Manish N Raizada, PhD
Department of Plant Agriculture
University of Guelph
Croplink Global Food Security Initiative
Guelph, ON N1G 2W1
Canada
Phone: +1 519 824 4120 ext 53396
Fax: +1 519 763 8933
Email: raizada@uoguelph.ca

Abstract

Background: While it is known that the majority of pages on the World Wide Web are in English, little is known about the preferred language of users searching for health information online.

Objectives: (1) To help global and domestic publishers, for example health and food agencies, to determine the need for translation of online information from English into local languages. (2) To help these agencies determine which language(s) they should select when publishing information online in target nations and for target subpopulations within nations.

Methods: To estimate the percentage of Web publishers that translate their health and food websites, we measured the frequency at which domain names retrieved by Google overlap for language translations of the same health-related search term. To quantify language choice of searchers from different countries, Google provided estimates of the rate at which its search engine was queried in six languages relative to English for the terms “avian flu,” “tuberculosis,” “schizophrenia,” and “maize” (corn) from January 2004 to April 2006. The estimate was based on a 20% sample of all Google queries from 227 nations.

Results: We estimate that 80%-90% of health- and food-related institutions do not translate their websites into multiple languages, even when the information concerns pandemic disease such as avian influenza. Although Internet users are often well-educated, there was a strong preference for searching for health and food information in the local language, rather than English. For “avian flu,” we found that only 1% of searches in non-English-speaking nations were in English, whereas for “tuberculosis” or “schizophrenia,” about 4%-40% of searches in non-English countries employed English. A subset of searches for health information presumably originating from immigrants occurred in their native tongue, not the language of the adopted country. However, Spanish-language online searches for “avian flu,” “schizophrenia,” and “maize/corn” in the United States occurred at only <1% of the English search rate, although the US online Hispanic population constitutes 12% of the total US online population. Sub-Saharan Africa and Bangladesh searches for health information occurred in unexpected languages, perhaps reflecting the presence of aid workers and the global migration of Internet users, respectively. In Latin America, indigenous-language search terms were often used rather than Spanish.

Conclusions: (1) Based on the strong preference for searching the Internet for health information in the local language, indigenous language, or immigrant language of origin, global and domestic health and food agencies should continue their efforts to translate their institutional websites into more languages. (2) We have provided linguistic online search pattern data to help health and food agencies better select languages for targeted website publishing.

(J Med Internet Res 2007;9(2):e18) doi:10.2196/jmir.9.2.e18

KEYWORDS

Health; Internet; Google; language; indigenous; food security; immigrant; avian flu; tuberculosis; maize; schizophrenia; nutrition; linguistic

Introduction

The World Wide Web has more than 15 billion pages [1] and has become an important source of health-related information [2-5]. In the United States, for example, it has been reported that 82% of female users and 77% of male users used the Internet to obtain medical information on a routine basis [6]. Google searches have even been shown to assist physicians in the correct diagnosis of medical ailments [7]. However, two thirds of the pages on the Web are published in English [8,9], even though the world has over 5 billion non-English speakers [10], including approximately 700 million non-English-speaking Internet users [8]. In fact, the vast majority of the world's 6900 living linguistic groups [10] have little Web content available in their language [8,9]. Adding to this problem, search engines such as Google do not translate search terms into other languages [9]—perhaps a surprise to many users. To overcome linguistic barriers, the World Health Organization (WHO) and the Food and Agricultural Organization (FAO) now publish their websites in six and four major languages, respectively. However, other globally authoritative organizations, such as the Centers for Disease Control (CDC) in the United States, primarily publish online information in only one or two languages of domestic importance.

In spite of the significant challenges created by linguistic differences in effectively communicating health information to the world's peoples online, we could find little quantitative data on this issue. Do the world's online users, presumably wealthier and more educated than the general population, primarily search online for health information in their local language, or do they employ Web-prevalent languages such as English? Are current online translation efforts by the world's health and food agencies beneficial, and should these agencies be spending more resources on these efforts? In a world of human migration, which language(s) should domestic governments and international agencies use in order to communicate online health information to target populations? To transmit information to front-line health professionals in developing nations, a group that can include international aid workers from wealthy nations, which language(s) should be employed to target a particular nation? Do indigenous peoples search for online health-related information using search terms belonging to their own language or the colonial language? Real-time, accurate communication of health information might be especially critical during a pandemic infectious disease outbreak or famine. To begin to answer these questions, we have used a case-study approach that examines linguistic preferences in Internet search engine queries.

Specifically, we measured search patterns on Google for four health- and food-related terms in seven languages in 227 nations. The four search terms we chose for our study were "avian flu," "tuberculosis," "schizophrenia," and "maize" (corn). We chose "avian flu" because it is an ideal model for searching for online information concerning an emerging infectious disease pandemic; as of August 2006, avian influenza (virus subtype H5N1) had killed 141 people in 10 countries [11,12] in addition to prompting the slaughter of millions of animals [13]. We chose "tuberculosis" because it is a good model for searching an established global infectious disease, as it is a major cause of death for HIV-positive patients [14] and currently afflicts about 15 million people in 207 nations [15]. We chose "schizophrenia" because psychiatric and neurological diseases affect more than 450 million people globally [16,17] and because online mental health information has the potential to help nations that have few mental health specialists [17]. Schizophrenia afflicts over 24 million people worldwide [18]. Finally, because there are currently around 850 million chronically undernourished people in the world [19], under frequent threat of famine, and because malnutrition is a major underlying contributor to infectious disease susceptibility [20], we chose "maize" (corn) because it is an important search term for global food security agencies. Maize supplies one third of all human calories in Latin America and Sub-Saharan Africa [19] and, combined with its genetic relatives rice and wheat, supplies approximately 50% of all human calories globally, either directly or via animal feed [19].

In this paper, we quantitatively demonstrate the need for health and food website translation that not only targets the world's major languages, but also linguistic minorities within nations, including immigrants, foreign aid workers, and indigenous groups.

Methods

Measuring the Extent to Which Health and Food Agencies Translate Websites

When an online publisher translates its Web site across languages, the same source domain name can be found in the different result lists of a search engine when entering different translated search terms. In an URL, the domain name is the identifier before the first slash (eg, for the WHO, the domain name is www.who.int). Therefore, to estimate the percentage of institutions, news agencies, and other sources that translate their health and food websites, we measured the frequency that domain names overlap (institutional sources overlap) in different results lists when Google is queried using different language translations of the same search term (eg, English "tuberculosis" versus Bahasa Indonesia "tuberkulosa"). The language translations used in this study are shown in Multimedia Appendix 1. We first extracted URLs from Google.com listed as "the most relevant" (about 500-1000 URLs) based on the Google page rank algorithm [21] on July 10, 2006. We then excluded nonunique domain names from within each query group; the remaining unique subset was then compared to the URLs retrieved using the comparison language.

Measurements of Online Search Rates and Language Choice

For each of the four terms and their translations, we measured the rate at which users from 227 different countries searched Google (Multimedia Appendix 2). The country of origin for each search was identified using the geographic locations of Internet Protocol (IP) addresses. The search rates were based on proprietary Google Inc. (Mountain View, CA) data using an algorithm that measured 20% of all first-page Web search queries from January 2004 to April 2006.

The search rate, *M*, for a term, *T*, originating from a specific country was then estimated from this large sample: the algorithm calculated the ratio of searches using a specific term (*T*) divided by all searches (*A*) from that nation during the time period, such that $M = T/A$. Multimedia Appendix 2 contains the standard error (SE) for each search rate. To calculate the standard error, a normal distribution for the statistic was assumed. To convert standard error to the 95% confidence interval of the estimate, the following formula was used: 95% CI = $M \pm (1.96 \times SE)$. Only the search rates for selected languages are shown for each nation, not the total search rate across languages or the total number of searches, which are proprietary.

For ease of database searching and validation, we chose languages that share a Latin-based script similar to English. For detailed analysis, the languages chosen were English, Bahasa Indonesia, Spanish, Portuguese, German, and French, as they represent the primary or secondary language of about 100-500 million people worldwide and/or are significant post-colonial languages [10]. Because Turkey has reported cases of avian influenza [11,12], Turkish was added as a language of interest. A search term may be common to multiple languages, not all of which are noted here. Multimedia Appendix 1 contains a complete list of the language translations used to retrieve search rate data from the Google database.

Health and Food Security Indicators

For the number of human avian influenza cases (subtype H5N1), we employed the WHO Epidemic and Pandemic Alert and Response Database [12], updated on August 9, 2006. For the number of poultry outbreaks of type H5 since 2003, we used the August 16, 2006 Avian Flu Update from the Organisation Mondiale de la Santé Animale (OiE) [World Organisation for Animal Health] [13]. For tuberculosis (TB), we used Millennium Development Goal (MDG) Indicator 23, the estimated TB prevalence in 2004 from the Global Tuberculosis Database of the WHO Global Health Atlas [15]. The number of psychiatrists per 100000 people in 2005 was obtained from Project Atlas: Resources for Mental Health and Neurological Disorders, from the WHO Global Health Atlas Database [17]. The most recent maize consumption data (kcal/person/day in 2004) was from the Supply Utilization Accounts (SUA) Database of the Food and Agricultural Organization Statistical Division (FAOSTAT) [19].

Results

Measuring the Extent to Which Institutions Currently Translate Online Information

We measured the extent to which the world's institutions, including health and food agencies, news organizations, and other sources, currently publish Web content in multiple languages. To quantify this, we measured the institutional source (host-domain) overlap between search results (first 500-1000 URLs) retrieved using English-language queries versus select comparison languages (Table 1). For example, as shown in Table 1, "avian flu" resulted in 906 hits, which included 539 unique hosts. The same search in French resulted in 801 hits, which included 375 hosts. Only 7.1% of the 539 hosts found in the English search were also found in the French search, while 10.1% of the 375 hosts found with the French search were also found in the English search. This means that approximately 7.1%-10.1% of all hosts have an English/French translation.

For "avian flu," which has afflicted 56 people in Indonesia [12], only 6.8% and 2.4% of institutional domain names overlapped between Bahasa Indonesia searches and English-language searches, respectively (Table 1). For "tuberculosis," we discovered only about 9% overlap in the institutional domain names retrieved for English and French/Portuguese translations: French and Portuguese are widely spoken by TB-afflicted nations in Africa and Brazil [15]. The host domain overlap was typically less than 10% between European languages; we were surprised by this low rate given the presence of common governmental institutions in Europe (eg, European Union). We extrapolate that 80%-90% of health- and food-related institutions do not translate their websites into multiple languages, even when the information concerns pandemic disease. This does not exclude the possibility that other agencies, such as domestic health agencies, might be translating this information.

We then specifically screened for Web pages belonging to the WHO and the CDC, both authoritative agencies for infectious disease information. Because the first page of search results is the most viewed [22], the rank order in which search results appear is critical. When we searched for "avian flu" in English on Google.com, the CDC website was the first hit, followed second by the WHO website. In the French translation, however, no CDC-affiliated Web pages appeared in the first 100 hits. When we searched Google Indonesia using the Bahasa Indonesia translation of "avian flu," a WHO-affiliated Web page did not appear until page three of the search results (rank 21), and the first CDC-affiliated page appeared on page six (rank 63). Most significant, by searching for "avian flu" in Turkish using Google Turkey, we were unable to retrieve the websites of either the WHO or CDC in the first 500 search results.

As a cautionary note, we found that a single accent or special character in the search term sometimes changed the rank order of health search results significantly, consistent with more detailed analysis conducted for nonmedical terms [9,23]: for example, for avian flu, we found that "gripe aviária" (accent, Portuguese) yielded 49/100 top search results of Brazilian origin (ie, "br" in domain name), whereas "gripe aviaria" (no accent, Portuguese) yielded 40/100 top search results from

Spanish-speaking nations and only 4/100 search results from Brazil. This is significant, because we found that Brazil, Portugal, and other Portuguese-speaking nations (Angola, Mozambique) searched Google for this term with and without the accent at nearly equal rates (Multimedia Appendix 2).

Table 1. Overlap in the institutional domain names retrieved by Google to measure the extent to which institutions translate websites across languages

English Search Term	Comparison Language	English Language Search			Comparative Language Search		
		URLs	Unique Hosts	Host Overlap with Comparison Languages	URLs	Unique Hosts	Host Overlap with English
Avian flu	French	906	539	7.1%	801	375	10.1%
Avian flu	Indonesian	906	539	2.4%	472	190	6.8%
Tuberculosis	French Portuguese Dutch	834	473	8.5%	790	426	9.4%
Tuberculosis	German Danish Afrikaans	834	473	0.0%	809	443	0.0%
Tuberculosis	Bahasa Indonesia	834	473	0.8%	462	273	1.5%
Schizophrenia	Spanish Portuguese	813	444	10.8%	764	518	9.3%
Schizophrenia	Bahasa Indonesia Malay (Bahasa Malayu)	813	444	17.1%	748	463	16.4%
Maize	Spanish	828	421	7.8%	830	543	6.1%

Language-Specific Searching of Infectious Disease Information

Though we did not retrieve many WHO- and CDC-affiliated Web pages when we searched across different languages, one possibility is that Internet users, many of whom are well-educated, are supplementing their Google searches for online health information by searching in English. If true, then there would be less of a need for the WHO and other global agencies to translate online information into diverse languages.

In our case study, we found the actual results to be variable (Table 2; Multimedia Appendix 2): for “avian flu,” we often found that only 1% of searches in non-English-speaking nations were in English, whereas for “tuberculosis” or “schizophrenia,” about 4%-40% of searches in non-English countries employed English. Brazil, which had an estimated 141000 cases of TB in 2004 [15], had an 18-fold higher query rate for “tuberculose” (Portuguese) than “tuberculosis” (English). However, the comparison is more important for languages that have much less Web content, such as Bahasa Indonesia and Turkish. Indonesia, which had the highest number of reported cases of human H5N1 viral infections in 2006 [12], had a 15-fold higher query rate for “avian flu” in Bahasa Indonesia (“flu burung”) than in English. We found that Turks used Turkish to search Google for health terms at 3- to over 1000-fold higher rates than English, French, German, or Spanish (Table 2; Multimedia Appendix 2). We did, however, find sites in Turkish that had translated information from the WHO. Whether or not the WHO, CDC, and FAO wish to leave it up to others to translate their information accurately and rapidly must be decided based on their confidence level of the eHealth capabilities of each target nation. Given the Google language-based search patterns, we conclude that during times of infectious disease outbreaks, though English may be useful, global agency–affiliated Web pages translated into local languages would likely be highly accessed and would have the benefits of being viewed as authoritative and accurate and of being transmitted in real time.

Language of Online Mental Health Information Searches

In terms of mental health, many developing nations have 10- to 100-fold fewer psychiatrists per capita than many developed nations [17] (Table 2). For this reason, global accessibility of online mental health information has the potential to be very beneficial to physicians and families of patients in developing nations [24]. Given ethnic taboos [24], we first asked whether or not Internet users from developing nations are searching online for mental health information—potentially useful information for global mental health experts. Indeed, we found that the search rate for “schizophrenia” was similar between developed and developing nations in the local language, demonstrating an active need for online mental health information in poor countries. We have made the full mental health dataset available (Multimedia Appendix 2). Given that Google estimates 3-fold more search results concerning schizophrenia in English versus the next 10 languages combined (data not shown), we asked whether middle- and low-income nations searched for this topic in English. We found that people from Brazil, a nation of about 180 million people, searched for “schizophrenia” in Portuguese at a 28-fold higher rate than in English (Table 2).

Table 2. Google search rates for selected health terms in local languages relative to English*

Search Term	Country	Public Health Comparison Metric	Local Language	English Searches (% of Local Language Searches)
Avian flu		**Human Cases† (Poultry Outbreaks)‡**		
	United States	0 (0)	English	100.0
	France	0 (1)	French	1.6
	Germany	0 (1)	German	0.8
	Turkey	12 (176)	Turkish	0.9
	Democratic Republic of Congo	0 (0)	French	1.1
	Cote d'Ivoire	0 (3)	French	1.4
	Burkina Faso	0 (4)	French	100.0
	Mozambique	0 (0)	Portuguese	143.8
	Mexico	0 (0)	Spanish	3.3
	Brazil	0 (0)	Portuguese	1.0
	Indonesia	56 (211)	Bahasa Indonesia	6.5
Tuberculosis		**TB Cases** §		
	United States	10510	English	100.0
	France	5901	French	9.8
	Germany	5243	German	14.3
	Turkey	32371	Turkish	29.1
	Democratic Republic of Congo	307554	French	23.6
	Cote d'Ivoire	116349	French	8.6
	Burkina Faso	46815	French	15.9
	Mozambique	123360	Portuguese	28.4
	Mexico	45710	Spanish	same term
	Brazil	141115	Portuguese	5.6
	Indonesia	605759	Bahasa Indonesia	1779.3
Schizophrenia		**Psychiatrists per 100000‖ People**		
	United States	13.70	English	100.0
	France	22.00	French	6.6
	Turkey	1.00	Turkish	10.5
	Democratic Republic of Congo	0.04	French	43.0
	Cote d'Ivoire	0.20	French	7.1
	Burkina Faso	0.05	French	9.5
	Mozambique	0.04	Portuguese	24.4
	Mexico	2.70	Spanish	4.1
	Brazil	4.80	Portuguese	3.6
	Indonesia	0.21	Bahasa Indonesia	115.4

*Based on sampling 20% of all searches on Google.com from January 2004 to April 2006.

†Subtype H5N1, from the WHO Epidemic and Pandemic Alert and Response Database [12], updated August 9, 2006.

‡Type H5 outbreaks since 2003, from the OiE [13], updated August 16, 2006.

[§]Estimated TB prevalence in 2004 from the WHO Global Tuberculosis Database [15].

[‖]2005 data from Project Atlas: Resources for Mental Health and Neurological Disorders, from the WHO Global Health Atlas Database [17].

In contrast, people from Indonesia, a nation of 220 million people, searched for "schizophrenia" at similar rates in Bahasa Indonesia as in English. As with infectious disease searching, our data would suggest that users from most nations tend to search Google for "schizophrenia" in their official language at up to 10-fold higher rates than other languages (Multimedia Appendix 2). Many of the world's people might therefore benefit if the world's most authoritative mental health agencies (eg, US National Institute of Mental Health) translated information into other languages, even though this is not part of their domestic mandate.

The Online Search Rates of Immigrant Minorities

In some developed nations, there is concern that immigrant groups might spread infectious diseases. In Europe, the TB prevalence is 43/100000 people in Turkey but lower in wealthier nations such as Germany (6/100000); in Asia, the TB prevalence is 273/100000 people in Indonesia, compared to 48/100000 in Singapore [15] (Table 2). Unlike many of their neighbors, both Turkey and Indonesia have reported human cases of avian influenza [12]. Governments may be interested to know whether their immigrant communities consult infectious disease–related websites originating from their adopted country or their native country. As a case study, we examined the search rates of Turks and Indonesians after they migrated to other nations in Europe or Asia, respectively, as determined by searches in Turkish and Bahasa Indonesia originating outside Turkey and Indonesia. The Turkish and Indonesian languages are distinct relative to many of their surrounding nations, permitting us to discern the search behavior of these populations after they have emigrated, provided they search in their native language. As shown in Table 3, we detected searches in Turkish for "avian flu" and "tuberculosis" throughout Europe, including high rates in Belgium and Austria, respectively. In Asia, we could detect searches in Bahasa Indonesia for "tuberculosis" and very high rates for "avian flu" throughout Asia and the Pacific region, including Hong Kong, Singapore, and Australia. We cannot exclude that these (presumed) Turkish and Indonesian immigrants also searched online in their adopted language(s), nor could we measure the fraction of their searches in their native language versus adopted language(s). However, based on the fraction of a country's population that belongs to a particular immigrant group versus the fraction of searches conducted in the immigrant language compared to the adopted language, then if every person in a country searched the Internet for the same term at the same rate, but in different languages, we could then extrapolate that Turkish immigrants in Belgium searched for "avian flu," "schizophrenia," and "tuberculosis" 100%, 40%, and 15% of the time, respectively, in Turkish rather than French (Table 3). Using the same simplistic assumptions, in Austria, 39% of searches for "tuberculosis" by Turkish immigrants were in Turkish rather than German. In reality, immigrant groups are likely searching for a term such as "avian flu" at a higher rate than the general population when the corresponding disease affects their homeland and is in the news. However, we also found high search rates for "tuberculosis" and even "schizophrenia" in Turkish in these nations, which are less featured in the news. We conclude that it is important for health officials to be aware that if they wish to disseminate health information to susceptible immigrant groups, they should not rely on websites published in the majority language(s) of the nation. High priority domestic health-related websites should be multilingual, particularly those that concern infectious disease.

Our analysis also revealed surprises: for example, in Bangladesh, one of the world's poorest and most populated nations, with 136 million residents, we found that the search rate for "avian flu" in Bahasa Indonesia ("flu burung") was equivalent to the search rate in English (Table 3); the Bahasa Indonesia translation of "schizophrenia" was also high in Bangladesh. The high Bahasa Indonesia search rate may reflect the fact that many Bangladeshi citizens work in Singapore and Malaysia, nations that speak a similar language, Malay (Bahasa Malayu) [25]; it would appear that when they return to Bangladesh, Bangladeshis continue to use the terminology they learned while away, but the reason is unclear. We suggest that international health organizations aiding Bangladesh should publish or meta tag health information in Malay (Bahasa Malayu) to reach health practitioners in that nation. This result also highlights the practical value of analyzing linguistic preferences during online searching.

Finally, we were surprised to find that online searches for "avian flu," "schizophrenia," and "maize/corn" in Spanish in the United States occurred at less than 1% of the English search rate (Table 3). The US online Hispanic population (legal and illegal) is estimated to be 12% of the total US online population, of which nearly half use Spanish for some (28%) or all (21%) of their Internet usage [8,26]. Therefore, our data suggest that Latin American immigrant groups in the United States search for health information to a lesser degree in Spanish than might be predicted, although these data could also be a result of the digital divide between the groups.

Table 3. Online search rates of immigrant minorities

Search Term	Country	Immigrant Language	Comparative Major Language* of Adopted Country	Immigrant Language Searches per 10000 Major Language Searches	Relevant Immigrants per 10000 Total Population†
Turkish immigrants in Europe					
Avian flu	Belgium	Turkish	French	56.1	39.7‡
	Switzerland	Turkish	German	9.2	N/A
	United Kingdom	Turkish	English	31.2	N/A
Tuberculosis	Austria	Turkish	German	60.2	155.4§
	Belgium	Turkish	French	5.8	39.7‡
	Germany	Turkish	German	11.8	212.2‖
	Switzerland	Turkish	German	10.7	N/A
	United Kingdom	Turkish	English	2.8	N/A
Schizophrenia	Belgium	Turkish	French	15.9	39.7‡
	Switzerland	Turkish	French	62.2	N/A
	United Kingdom	Turkish	English	2.9	N/A
Indonesian/Malaysian immigrants in Asia/Pacific					
Avian flu	Australia	Bahasa Indonesia	English	234.7	625.2¶
	Bangladesh	Bahasa Indonesia	English	10474.0	N/A
	Hong Kong	Bahasa Indonesia	English	3424.4	N/A
	India	Bahasa Indonesia	English	154.4	N/A
	Singapore	Bahasa Indonesia	English	2268.0	N/A
Tuberculosis	Hong Kong	Bahasa Indonesia	English	23.3	N/A
	Singapore	Bahasa Indonesia	English	17.0	N/A
Schizophrenia	Australia	Bahasa Indonesia	English	6.0	625.2¶
	Bangladesh	Bahasa Indonesia	English	687.4	N/A
	Hong Kong	Bahasa Indonesia	English	295.1	N/A
	India	Bahasa Indonesia	English	16.0	N/A
Spanish immigrants in English-Speaking North America#					
Avian flu	Unites States	Spanish	English	102.8	> 503.0**
Schizophrenia	Unites States	Spanish Portuguese	English	91.8	> 503.0**
Corn (maize)‡‡	Unites States	Spanish Indigenous (“choclo”/”elote”)	English	37.9	> 503.0**

*a major language for which Google search rates were accessible, not necessarily the largest linguistic group of the nation.

†Total population data are from United Nations Population Division, 2005 data. URL: http://www.un.org/esa/population/unpop.htm.

‡Data are from Institut National De Statistique, 2003 data. Population et Ménages: Mouvement de la population et migrations, "Immigrations extérieures par nationalité et groupe d'âges – Belgique." URL: http://www.statbel.fgov.be.

§Data are from Statistics Austria, Volkszählung. Hauptergebnisse I – Österreich, 2001 census data. URL: statistik.at/neuerscheinungen/vzaustria.shtml.

‖Data are from Statistisches Bundesamt (Federal Statistical Office, Germany), 2006 data. URL: http://www.destatis.de/themen/e/thm_bevoelk.htm.

¶Data are from Australia Bureau of Statistics, Cultural and Language Diversity, 2001 Census data. URL: http://www.abs.gov.au/.

#“Tuberculosis” is not included as it is the same term in both English and Spanish.

**Data are from Department of Homeland Security Yearbook of Immigration Statistics. Legal immigrants, 2005 data. URL: uscis.gov/graphics/shared/aboutus/statistics/ybpage.htm.

‡‡The search term used was “corn.”

Table 4. Search rates in European languages in Sub-Saharan Africa

Country	Colonial Language*	Minority Language	Minority Language Searches per 10000 Colonial Language Searches
Search Term: Avian Flu			
Angola	Portuguese	English	6629
Cameroon	French	German	323
Ghana	English	French	2024
Ghana	English	German	1744
Ghana	English	Dutch	498
Kenya	English	French	2941
Mozambique	Portuguese	English	14286
Mozambique	Portuguese	French	4444
Nigeria	English	French	4235
Nigeria	English	German	897
Rwanda	French	English	1481
Senegal	French	English	732
Search Term: Tuberculosis			
Angola	Portuguese French	English Spanish	1504
Cameroon	French Portuguese	English	2405
Democratic Republic of Congo	French Portuguese	English Spanish	2364
Ghana	English	French Portuguese	836
Kenya	English	French Portuguese	1128
Mozambique	Portuguese French	English Spanish	2842
Nigeria	English	French Portuguese	506
Rwanda	French Portuguese	English Spanish	2808
South Africa	English	Afrikaans	1547
Search Term: Schizophrenia			
Angola	Portuguese Spanish	English	1547
Cameroon	French	English	1922
Democratic Republic of Congo	French	English	4304
Ghana	English	French	488
Mozambique	Portuguese Spanish	English	2441
Rwanda	French	English	18453
Senegal	French	English	3287

*When two languages are noted, the term is the same in both languages.

Search Rates in European Languages in Sub-Saharan Africa

Sub-Saharan Africa suffers from high rates of infectious disease, including TB and HIV [18], and high rates of malnutrition [19]. As demonstrated in Table 4, our search results suggest that the dissemination of online health or food security information to this region by international agencies should not be limited to the colonial language(s) of the target nations. For example, in English-speaking Ghana, 20% of searches for avian flu were in French ("grippe aviaire"), 17% in German ("vogelgrippe"), and 5% in Dutch ("vogelgriep") relative to English (Table 4). In Mozambique, a former Portuguese colony, we found 1.4 times more searches for "avian flu" in English relative to Portuguese, and a high rate in French. In the Democratic Republic of Congo (DRC), a former French colony with more than 300000 TB infections in 2004 [15], 24% of the searches for tuberculosis were in English relative to French. Because Internet use is only 1.9% is Ghana, 0.7% in Mozambique, and 0.2% in DRC [8], it is plausible that these high rates of non-colonial language searches using Google may reflect searches by health professionals trained in other nations, including workers from international agencies who would be expected to have better Internet access than the general population.

The Effect of Region-Specific Cultural and Indigenous Terminology

Finally, we measured the effects of cultural bias within the same linguistic group, using a term important for human nutrition and food security, "maize." Maize is eaten directly, but it is also a major source of animal feed worldwide. Maize is known as "corn" in the United States and the United Kingdom, but as "maize" in many other English-speaking nations. We found that the US searches for "corn" were at a 28-fold higher rate than for "maize," while other English-speaking nations such as Nigeria and Zimbabwe queried "maize" at a 1.5- to 4-fold higher rate than "corn," respectively (Table 5); in the latter nations, "corn" may also refer to any large cereal grain (eg, wheat). This search behavior has consequences, as we found only three domain names that overlapped between searches for "maize" versus "corn" out of the first 50 unique Google search results.

Therefore, when African nations attempt to retrieve information about growing maize in English, they may unknowingly be excluding authoritative information from the United States and other Western English-speaking nations, such as practical information from the US National Corn Growers' Association, whose website does not appear in the first 100 Google hits for "maize," but ranks sixth when "corn" is searched. Similarly, when international organizations such as the FAO wish to transmit knowledge, for example, using Spanish or Portuguese in Latin America, indigenous terminology usage may make some of this information inaccessible. Maize originated from Southern Mexico and fed indigenous Latin American civilizations [27]. In Mexico and Guatemala, we found that an indigenous synonym for maize, "elote," was searched at a similar rate to the Spanish term, "maíz" (Table 5). In Peru, however, we found that a different indigenous term, "choclo," was searched at a 2-fold higher rate than "maíz." We conclude that cultural and indigenous linguistic divisions may be preventing large numbers of food security and nutrition websites from reaching those people aiding 800 malnourished people or 1.2 billion agricultural workers that live in developing countries [19]. Cultural and indigenous bias may be particularly prevalent for terms related to crops, diseases, or pathogens that have pre-colonial origins.

Table 5. The effect of region-specific cultural and indigenous terminology

Country	Calories from Maize (kcal/person/day)*	Search Term Comparison	Search Rate Ratio
Canada	N/A†	corn:maize	36:1
United States	512	corn:maize	28:1
United Kingdom	115	corn:maize	7:1
India	38	corn:maize	2:1
Nigeria	179	corn:maize	1:2
Kenya	775	corn:maize	1:3
Tanzania	646	corn:maize	1:4
Zimbabwe	720	corn:maize	1:4
Spain	N/A†	maíz:elote:choclo	10:1:4
Venezuela	467	maíz:elote:choclo	16:1:3
Colombia	312	maíz:elote:choclo	20:1:5
Mexico	1081	maíz:elote:choclo	14:16:1
Guatemala	869	maíz:elote:choclo	12:10:1
Peru	145	maíz:elote:choclo	11:1:25
Argentina	132	maíz:elote:choclo	41:1:51

*Data from FAOSTAT [19].

†N/A: Data not available.

Discussion

Principal Results

In a world where infectious disease pandemics and threats of famine are always present, and in spite of the fact that the World Wide Web offers great hope for rapid and accurate sharing of information between peoples, we have demonstrated that one linguistic group does not or cannot access the health and food security websites of a different linguistic group. Our data suggest at least three reasons for this.

The first reason is that the websites of most institutions are not published in more than one or two languages. When we sampled Web pages found by the English-language queries "avian flu," "tuberculosis," "schizophrenia," and "maize/corn" and their counterpart queries in other languages, the Google search results typically overlapped by only less than 10% in terms of the domain names retrieved, indicating that 90% of the relevant Web pages had not been translated into at least two languages (see Table 1). For example, when Turkish or Bahasa Indonesia was used as the search language, Web pages from very authoritative sources, such as the CDC or WHO, were not retrieved by Google. We also found that a single linguistic accent or special character in the search query could significantly alter the number and content of health-related search results retrieved by Google. Therefore, one reason for the linguistic digital divide is that the majority of health and food Web pages are not translated into multiple languages and/or that their cross-language retrieval by search engines is poor.

The first problem would not be important if the world's online community, better educated than the general public, searched in English, since the majority of the world's Web pages are published in English [8,9]. However, we found that there was a 2- to 100-fold higher Google search rate for health and food terms in the native language of a country compared to English (see Table 2).

Finally, within a nation, it might be assumed that language would not be a problem if a domestic agency only published their health websites in the majority language of their own people. However, we found that in Asia and Western Europe, a subset of immigrants from Indonesia and Turkey, respectively, searched Google for health and food information in their native language, not the language(s) of their adopted countries (see Table 3). In Sub-Saharan Africa, we detected unexpectedly high search rates for health information in non-colonial European languages (see Table 4), perhaps reflecting the presence of international aid workers. Finally, in Latin America, we found that indigenous words were used to search Google for information about food, rather than the colonial language of Spanish (see Table 5). Therefore, domestic agencies, in addition to global agencies, face a linguistic challenge when publishing information online: their target audiences still require information to be published in different languages, even though Internet users are presumably more educated and thus more multilingual than the general population.

Recommendations

Given our observation that the world's peoples appear to be searching for health and food terms in their local language or mother tongue, not in English, previous online language translation efforts by the WHO and FAO have no doubt been worthwhile. This is also revealed by the high page rank (first page results) of WHO-affiliated search results when Google is searched for infectious disease information in one of the WHO's six online languages. We recommend that these efforts be continued and further expanded to include more languages; this recommendation applies to global agencies, but also to domestic agencies, in order to meet immigrant or indigenous needs and/or to make information accessible to other nations. To achieve such translation goals, better health-specific translation software must be developed and more translators are needed who specialize in human health and food security terminology. For example, improved cross-language search retrieval [23] of health information by online search engines would be beneficial. These investments should then be used to initially target health and food security information to the world's most important linguistic groups, which include speakers of Chinese (1080 million people), Hindi (about 500 million), English (350-500 million), Spanish (390 million), Arabic (255 million), Portuguese (190 million), Bengali (215 million), Russian (255 million), Bahasa Indonesia (200 million), Japanese (127 million), Punjabi (104 million), German (123 million), and French (119 million) [10]. In a world that is primarily non-English speaking, such attempts will help to reduce the linguistic digital divide in health and food information on the World Wide Web.

Furthermore, as we did in this study (see Table 2 to Table 5), we recommend that when global or domestic health and food security organizations wish to use the Internet to disseminate information to other nations [28] or to their own immigrant or indigenous communities, they should first consult search engine query rates for different translations of possible search terms in order to determine which online languages are most needed. Multimedia Appendix 2 contains extensive linguistic online search pattern data to help health and food agencies better select languages for targeted website publishing. In order to measure search rates for other subjects of interest, we note that free online tools exist, such as Google Trends. In some situations, such as targeting indigenous groups, who often speak the majority language of a nation, all that may be needed is to imbed translated keywords into a majority-language website (eg, Spanish) so that search engines such as Google can cross-retrieve relevant information.

Future Studies

Though this study examined the extent to which agencies such as the WHO are publishing information in multiple languages, we did not systematically address the quality of health and food information available in different languages. Quality analysis would especially be important for minor languages that have little content available on the Web: for example, we estimate that there are more than 4000-fold fewer search results in Bahasa Indonesia than in English for "tuberculosis," more than 200-fold fewer search results for "avian flu" in Arabic or Japanese than

in English, and about 500-fold fewer search results for "schizophrenia" in Arabic than in English. Though these precise numbers are considered to be unreliable [29], they do illustrate the point that most of the world's linguistic groups, even major ones, have much less available online health information relative to English. As to the quality of the websites that are available, this will require systematic analysis, which poses significant methodological problems [30]. One could, however, perform a subjective survey-based evaluation by multilingual physicians, as has been performed to evaluate disease-specific websites published in English [31]. Such a survey should include quantifying to what extent online information from the WHO, CDC, and FAO is being translated into the world's minor languages in order to help these agencies determine where they need to target their online translation efforts.

Acknowledgments

This work was funded by an Ontario Premier's Research Excellence Award to MNR as part of the CropLink Online Global Initiative. No external sponsors were involved in the preparation or review of this manuscript.

Conflicts of Interest

Olcan Sercinoglu and Artem Boytsov are employees of Google Inc.

Multimedia Appendix 1

Language translations used in this study (pdf) [PDF file (Adobe Acrobat), 32 KB - jmir_v9i2e18_app1.pdf]

Multimedia Appendix 2

Rate at which users from 227 different countries searched Google (xls) [XLS file (MS Excel), 211 KB - jmir_v9i2e18_app2.xls]

References

1. Pandia Search Engine News. The size of the World Wide Web. 2007 Feb 25. Pandia.com. URL: http://www.pandia.com/sew/383-web-size.html [accessed 2007 Jun 14] [WebCite Cache ID 5NWELVyMw]
2. Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer Internet search activity on a major search engine, United States 2001-2003. J Med Internet Res 2005 Jul 1;7(3):e36 [FREE Full text] [Medline: 15998627] [doi: 10.2196/jmir.7.3.e36]
3. Phillipov G, Phillips PJ. Frequency of health-related search terms on the Internet. JAMA 2003 Nov 5;290(17):2258-2259. [Medline: 14600183] [doi: 10.1001/jama.290.17.2258]
4. Eysenbach G, Köhler C. Health-related searches on the Internet. JAMA 2004 Jun 23;291(24):2946. [Medline: 15213205] [doi: 10.1001/jama.291.24.2946]
5. Giustini D. How Google is changing medicine. BMJ 2005 Dec 24;331(7531):1487-1488 [FREE Full text] [Medline: 16373722] [doi: 10.1136/bmj.331.7531.1487]
6. Fox S. Online Health Search 2006: Most internet users start at a search engine when looking for health information online. Very few check the source and date of the information they find. Washington, DC: Pew Internet & American Life Project; 2006. URL: http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf [accessed 2007 Jan 18] [WebCite Cache ID 5LzT43f7w]
7. Tang H, Ng JHK. Googling for a diagnosis--use of Google as a diagnostic aid: internet based study. BMJ 2006 Dec 2;333(7579):1143-1145 [FREE Full text] [Medline: 17098763] [doi: 10.1136/bmj.39003.640567.AE]
8. Gordon R. Ethnologue: Languages of the World. 15th edition. Dallas, TX: SIL International; 2005. URL: http://www.ethnologue.com/ [accessed 2007 Jun 14]
9. Usage and Population Statistics. 2006. Internet World Stats. URL: http://www.internetworldstats.com/ [accessed 2007 Jun 14] [WebCite Cache ID 5Koa9heck]
10. Bar-Ilan J, Gutman T. How do search engines respond to some non-English queries? J Info Sci 2005;31(1):13-28. [doi: 10.1177/0165551505049255]
11. Epidemiology of WHO-confirmed human cases of avian influenza A(H5N1) infection. Wkly Epidemiol Rec 2006 Jun 30;81(26):249-257 [FREE Full text] [Medline: 16812929]
12. Epidemic and Pandemic Alert and Response Database. World Health Organization (WHO). Geneva, Switzerland: World Health Organization; 2006. URL: http://www.who.int/csr/en/ [accessed 2006 Aug 9] [WebCite Cache ID 5KoaCRAcS]
13. ; Organisation Mondiale de la Santé Animale (OiE) [World Organisation for Animal Health]. Animal Diseases Data: Avian Influenza. Paris, France: World Organisation for Animal Health; 2006. URL: http://www.oie.int/eng/info/hebdo/A_DSUM.htm [accessed 2006 Aug 16] [WebCite Cache ID 5KoannDMB]

14. Harries AD, Dye C. Tuberculosis. Ann Trop Med Parasitol 2006;100(5-6):415-431. [Medline: 16899146] [doi: 10.1179/136485906X91477]
15. ; World Health Organization (WHO). Global Tuberculosis Database: Global Health Atlas. Geneva, Switzerland: World Health Organization; 2004. URL: http://www.who.int/globalatlas [accessed 2006 Aug 24] [WebCite Cache ID 5Kob9nRsj]
16. ; World Health Organization (WHO). The World Health Report: 2001: Mental Health: New Understanding, New Hope. Geneva, Switzerland: World Health Organization; 2001. URL: http://www.who.int/entity/whr/2001/en/whr01_en.pdf [WebCite Cache ID 5Ps2dJIjc]
17. ; World Health Organization (WHO). Project Atlas: Resources for Mental Health and Neurological Disorders. Geneva, Switzerland: World Health Organization; 2005. URL: http://globalatlas.who.int/ [accessed 2006 Aug 28] [WebCite Cache ID 5KobJwBPx]
18. Mathers C, Stein C, Ma FD, Rao C, Inoue M, Tomijima N. Global Burden of Disease 2000: Version 2 Methods and Results. GPE Discussion paper No. 50. Geneva, Switzerland: World Health Organization; 2002. URL: http://www.who.int/entity/healthinfo/paper50.pdf [WebCite Cache ID 5Pw4lFVxS]
19. ; Food and Agricultural Organization, Statistical Division (FAOSTAT). FAOSTAT Database. Rome, Italy: FAOSTAT; 2006. URL: http://faostat.fao.org/ [accessed 2006 Apr 26] [WebCite Cache ID 5KobXBoYC]
20. Keusch GT. The history of nutrition: malnutrition, infection and immunity. J Nutr 2003 Jan;133(1):336S-340S [FREE Full text] [Medline: 12514322]
21. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. Comput Networks ISDN 1998;30(1-7):107-117. [doi: 10.1016/S0169-7552(98)00110-X]
22. Wolfram D, Spink A, Jansen B, Saracevic T. Vox populi: the public searching of the Web. J Amer Soc Info Sci 2001;52(12):1073-1074. [doi: 10.1002/asi.1157]
23. Freitas-Junior H, Ribeiro-Neto B, Vale R, Laender A, Lima L. Categorization-driven cross-language retrieval of medical information. J Amer Soc Info Sci Tech 2006;57(4):501-510. [doi: 10.1002/asi.20320]
24. Miller G. Mental health in developing countries. The unseen: mental illness's global toll. Science 2006 Jan 27;311(5760):458-461. [Medline: 16439637] [doi: 10.1126/science.311.5760.458]
25. Rahman MM. Emigration and development: the case of a Bangladeshi village. Int Migr 2000;38(4):109-130. [Medline: 12349805] [doi: 10.1111/1468-2435.00122]
26. A Year in Review: U.S. Hispanic Internet Usage. ComScore Press Release, Comscore Networks. 2003 Nov 13. URL: http://www.comscore.com/press/release.asp?press=370 [accessed 2006 Oct 5] [WebCite Cache ID 5KobZKZHe]
27. Sluyter A, Dominguez G. Early maize (Zea mays L.) cultivation in Mexico: dating sedimentary pollen records and its implications. Proc Natl Acad Sci U S A 2006 Jan 24;103(4):1147-1151 [FREE Full text] [Medline: 16418287] [doi: 10.1073/pnas.0510473103]
28. Edejer TT. Disseminating health information in developing countries: the role of the internet. BMJ 2000 Sep 30;321(7264):797-800 [FREE Full text] [Medline: 11009519] [doi: 10.1136/bmj.321.7264.797]
29. Bar-Ilan J. Expectations versus reality – Search engine features needed for Web research at mid 2005. Cybermetrics 2005;9:Paper 2.
30. 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]
31. Craigie M, Loader B, Burrows R, Muncer S. Reliability of health information on the Internet: an examination of experts' ratings. J Med Internet Res 2002 Jan 17;4(1):e2 [FREE Full text] [Medline: 11956034] [doi: 10.2196/jmir.4.1.e2]

Abbreviations

CDC: Centers for Disease Control
DRC: Democratic Republic of Congo
FAO: Food and Agricultural Organization
MDG: Millennium Development Goal
OiE: Organisation Mondiale de la Santé Animale [World Organisation for Animal Health]
TB: tuberculosis
WHO: World Health Organization

submitted 01.12.06; peer-reviewed by J Bar-Ilan; comments to author 22.12.06; revised version received 23.03.07; accepted 17.05.07; published 28.06.07.

Please cite as:
Singh PM, Wight CA, Sercinoglu O, Wilson DC, Boytsov A, Raizada MN
Language Preferences on Websites and in Google Searches for Human Health and Food Information
J Med Internet Res 2007;9(2):e18
URL: http://www.jmir.org/2007/2/e18/
doi:10.2196/jmir.9.2.e18
PMID:17613488

Original Paper

How Cancer Survivors Provide Support on Cancer-Related Internet Mailing Lists

Andrea Meier[1], PhD; Elizabeth J Lyons[2], MPH; Gilles Frydman[3]; Michael Forlenza[4], MPH, PhD; Barbara K Rimer[2], DrPH

[1]School of Social Work, The University of North Carolina, Chapel Hill, NC, USA

[2]School of Public Health, Department of Health Behavior and Health Education, The University of North Carolina, Chapel Hill, NC, USA

[3]Association of Cancer Online Resources (ACOR), New York, NY, USA

[4]Health Sciences Department, Simon Fraser University, Vancouver, BC, Canada

Corresponding Author:
Andrea Meier, PhD
School of Social Work
The University of North Carolina at Chapel Hill
301 Pittsboro St
Chapel Hill, NC 27599-3550
USA
Phone: +1 919 962 4366
Fax: +1 919 843 8715
Email: ameier@email.unc.edu

Abstract

Background: Internet mailing lists are an important and increasingly common way for cancer survivors to find information and support. Most studies of these mailing lists have investigated lists dedicated to one type of cancer, most often breast cancer. Little is known about whether the lessons learned from experiences with breast cancer lists apply to other cancers.

Objectives: The aim of the study was to compare the structural characteristics of 10 Internet cancer-related mailing lists and identify the processes by which cancer survivors provide support.

Methods: We studied a systematic 9% sample of email messages sent over five months to 10 cancer mailing lists hosted by the Association of Cancer Online Resources (ACOR). Content analyses were used to compare the structural characteristics of the lists, including participation rates and members' identities as survivors or caregivers. We used thematic analyses to examine the types of support that list members provided through their message texts.

Results: Content analyses showed that characteristics of list members and subscriber participation rates varied across the lists. Thematic analyses revealed very little "off topic" discussion. Feedback from listowners indicated that they actively modeled appropriate communication on their lists and worked to keep discussions civil and focused. In all lists, members offered support much more frequently than they requested it; survivors were somewhat more likely than caregivers to offer rather than to ask for support. The most common topics in survivors' messages were about treatment information and how to communicate with health care providers. Although expressions of emotional support were less common than informational support, they appeared in all lists. Many messages that contained narratives of illness or treatment did not specifically ask for help but provided emotional support by reassuring listmates that they were not alone in their struggles with cancer. Survivors' explicit expressions of emotional support tended to be messages that encouraged active coping. Such messages also provided senders with opportunities to assume personally empowering "helper" roles that supported self-esteem.

Conclusions: Many cancer survivors use the Internet to seek informational and emotional support. Across 10 lists for different cancers, informational support was the main communication style. Our finding of an emphasis on informational support is in contrast to most prior literature, which has focused on emotional support. We found the most common expressions of support were offers of technical information and explicit advice about how to communicate with health care providers. Topics and proportions of informational and emotional support differed across the lists. Our previous surveys of ACOR subscribers showed that they join the lists primarily to seek information; this qualitative study shows that they can and do find what they seek. They also find opportunities to play rewarding roles as support givers.

(J Med Internet Res 2007;9(2):e12) doi:10.2196/jmir.9.2.e12

KEYWORDS

Internet; cancer; patients; survivors; online communities; mailing lists; online support groups; qualitative research

Introduction

The importance of the Internet as a health resource is demonstrated by the fact that 8 out of 10 Internet users in 2005 reported looking for health information online, most commonly seeking information on specific diseases and certain medical treatments [1]. In the United States, it is estimated that 56.3 million people actively seek information about chronic diseases [2], and the information they gather affects their health choices [3, 4]. Internet users report employing the information they find on the Internet to diagnose health problems, enhance their medical care, and to validate the advice they receive from doctors [5].

An estimated 100 million Americans report ever having been members of some type of online group, and 79 million Americans have become members of online support groups [6]. On the popular Yahoo site alone, users can choose from over 30000 health-related support groups. The World Wide Web and email now permit a variety of group communication formats, many of which are widely used and have been described in detail elsewhere [7]. Here, we will describe our work with one type of group format—the mailing list, also known as an email discussion group. In Internet mailing lists, email messages (asynchronous communication) from authorized senders (subscribers) may convey information and support to other list subscribers. In the case of eHealth support lists, many subscribers are living with similar health conditions or are caregivers to survivors.

Some sources estimate that as many as 1 in 4 disease information seekers join online discussion groups [8]. Approximately 23 million people are reported as very active in online communities [9]. Our recent count of 33000 health-related online self-help groups on Yahoo shows that participation in electronic support groups (ESGs) continues to grow. This estimate represents a 32% increase over the number reported by Eysenbach and colleagues in 2004 [10]. Although estimates vary greatly, millions of people in the United States and, increasingly, around the world are turning to online support groups to deal with health concerns. (Most online support lists are hosted in the United States but are accessible outside US boundaries. For a description of online support groups sited outside the United States, see [11]).

ESGs Within Virtual Social Networks

There is an ongoing debate about whether support lists should be considered ESGs, informal grass roots virtual organizations, or electronically networked communities.

Mailing Lists as Support Groups

Support mailing lists are similar to traditional offline self-help groups in that they are "composed of members who share a common condition, situation, heritage, symptom or experience [12]." eHealth support lists and offline self-help groups share the goal of helping people learn about and cope with a variety of risk factors, diseases, and conditions.

Typically, offline face-to-face support groups are small, composed of 10 to 12 members. In face-to-face groups, the small size makes it easier for members to interact with each other, to build trusting relationships, and for the groups to become cohesive [13]. By contrast, online mailing lists can have hundreds or even thousands of members, many of whom post messages infrequently, if ever. Using a liberal definition of participation—at least one post within a three-month period—Nonnecke and Preece found that only about 55% of subscribers to a virtual health support group could be described as active participants [14]. Those members who post with some regularity often become acquainted and emotionally bonded with each other, forming subgroups that function like cohesive face-to-face support groups. The impact of participation on lurkers, those who read messages but don't write them, is unknown.

Currently, most ESGs appear not to be professionally facilitated but rely on peer leaders, making them more like self-help or mutual aid groups than professionally facilitated face-to-face support group interventions. The Association of Cancer Online Resources (ACOR) mailing lists we studied follow the peer leader model. Many of the listowners are extremely knowledgeable about health and cancer. These peer-leader listowners intervene both online and offline as needed to correct misconceptions, enforce group norms, and provide information, but they aim to do so as infrequently as possible [12].

Support Lists as Grassroots Organizations and Virtual Communities

The list subgroups described above exist concurrently within larger networks that resemble grassroots organizations. Because mailing lists are embedded in the Web, members can follow links to additional information sources. In many ways, lists act as a portal for members, leading them to further explore the Web and to discover the external, socioeconomic, and structural factors that contribute to their health concerns. In this respect, support lists are like their offline grassroots counterparts in that they can organize around "communities of interest" to address social injustice. (In this instance, the injustice involves the unmet needs for support, access to treatment, and resources for cancer-affected people [15, 16]). As in offline grassroots organizations, support lists tend to be composed of members who share similar concerns, and the groups' informal organizational structures enable quick response to changing circumstances.

Research on Therapeutic Factors and Outcomes of ESG Participation

Anecdotal and descriptive information about online self-help processes suggest that virtual communities are possibly the most important aspect of the Web, with the biggest impact on health outcomes [10]. Research in this area is still in its early development; consequently, rigorous studies documenting these benefits are difficult to find. Much of the research to date has focused on describing how social support is communicated online [17,18] or how Internet communication has made it

possible to offer support to greater numbers of people—especially those with rare diseases [19,20]—in ways that are satisfying and empowering to most participants [18,21-23]. Because ESG participants are invisible to each other, it is easier for members to communicate about common concerns. Participants, particularly patients in illness support groups, do not have to be concerned whether their personal appearance will affect others' reactions to them, and race, gender, and other sociodemographic differences are not immediately apparent [24]. Members may increase their self-confidence by becoming better informed about their illnesses. These processes appear to enhance self-esteem and increase participants' comfort level in dealing with health professionals [5]. Participation in ESGs may help cancer survivors find information, obtain support, formulate questions to ask health care providers, and become more active partners in their care decisions [25]. However, while prior reports are encouraging regarding the impact of ESGs, the data were from uncontrolled studies.

Research Methodology and ESGs

Researchers have found naturally forming online groups that offer peer-to-peer social support difficult to study using conventional methods because both format and content are difficult to replicate using controlled methods [10]. Some have tried to cope with this methodological challenge by forming project-specific ESGs as components of multi-modality intervention studies. Typically, these ESGs have used closed memberships, trained facilitators, and limited brief durations. In this respect, such project-specific ESGs used as formal interventions have greater similarity to face-to-face support groups [13].

In a recent systematic review of 38 studies on the effects of peer-to-peer interactions in health-related virtual communities and ESGs, Eysenbach and colleagues concluded that only six studies evaluated pure peer-to-peer communities [10]. In the Eysenbach review, qualifying studies tended to have "less than optimum research designs" in that they were exploratory in nature, used nonexperimental designs, and had small sample sizes. One study had a 2 × 2 factorial design (full version website or control group website combined with or without peer-to-peer groups) that allowed an evaluation of the peer-to-peer component; the 31 remaining studies evaluated complex interventions in which online communities were only an adjunct to broader interventions [10]. These findings were similar to those of an earlier systematic review of the research on online cancer support groups (published from 1993-2002) conducted by Klemm and colleagues [17].

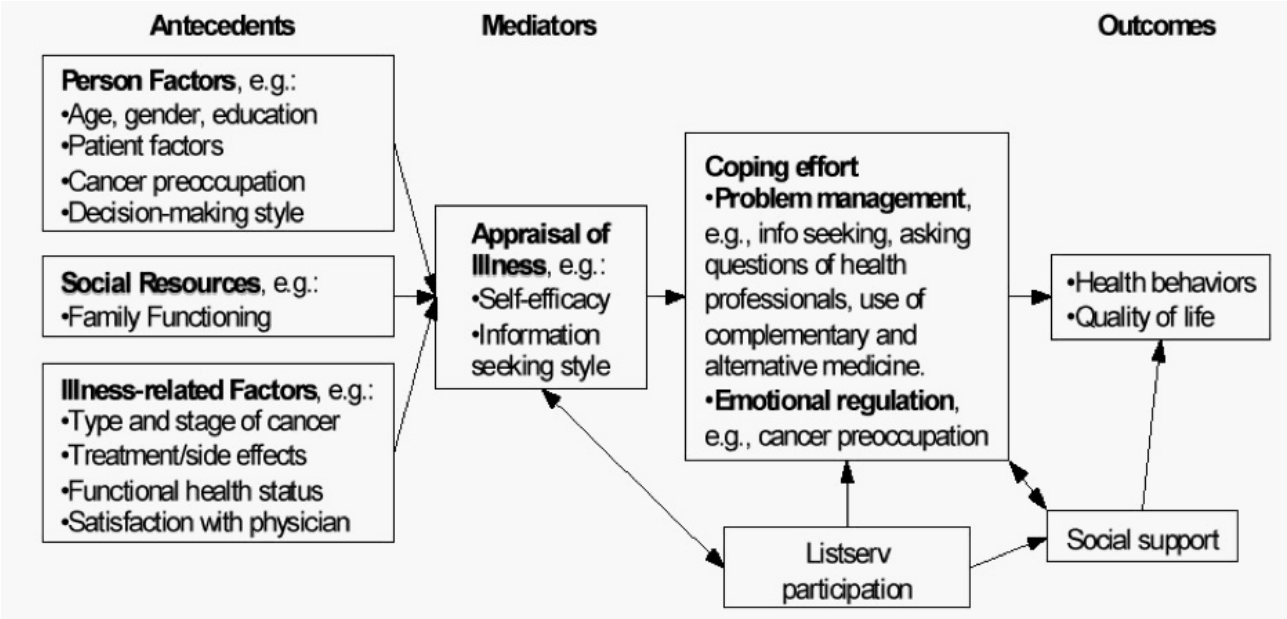
In the six peer-to-peer community studies included in the Eysenbach review, the type of ESGs varied across studies and included Web-based discussion forums, chat groups, combinations of chat and newsgroups, mailing lists, and one voice bulletin board system. All ESGs included in these studies had some degree of formal facilitation by health professionals. The role of health professionals as facilitators was to stimulate discussions by posing questions to the group, to post topics of interest to the group, or to provide educational materials. Some studies that used project-specific ESGs observed a possible dose-response effect between higher rates of participation and better outcomes for problems such as depression [26,27], caregiver strain [28], and increased perceived social support among people with diabetes [29]; however, the direction of causality is uncertain. Given the dearth of research in this field, we can only affirm that findings about the benefits of ESGs are promising but inconclusive. It is worth mentioning that Eysenbach and colleagues noted that no negative effects were reported. They further concluded that, because of the complexity of ESGs and methodological challenges, lack of evidence should not lead to the conclusion that ESGs are ineffective [10]. Rather, there is insufficient evidence regarding their efficacy. These cumulative findings suggest that ESGs—both pure peer-to-peer and short-term interventions—may play important mediating and moderating roles, creating the conditions that promote participant self-efficacy and positive health behaviors [10,29]. This research focuses on the extent to which peer-to-peer groups for different types of cancer contribute to these constructive attitudes and behaviors.

Conceptual Framework

The underlying conceptual framework for the Health eCommunities Project (for detailed description see [30]) was applied in developing the initial coding framework for our analysis. Briefly, the framework is informed by a general model of stress and coping [31-33] based on Lazarus and Folkman's theory of stress and coping [34] (Figure 1). Coping is a process of managing stress, initiated when a person appraises problems as exceeding their individual resources [35]. Here, we use two major categories of coping: problem-focused and emotion-focused [34]. Problem-focused coping entails constructive action to change the stressful situation. In the context of online mailing lists, problem-focused coping might be seeking treatment information online or researching complementary and alternative medicine. Emotion-focused coping uses actions to change an individual's emotional response to the stressful situation. Emotion-focused coping in an online community might occur when a member vents strong emotions by writing about them or when he or she avoids or denies such feelings.

Either coping strategy can lead to positive or negative outcomes depending on how well matched it is to the situation. According to this framework, when people confront stress, personality characteristics, external resources, and social support can influence coping, thereby mediating the effect on psychological outcome [36]. Social support can facilitate an individual's positive efforts to cope, potentially bolstering both positive problem-focused and emotion-focused coping.

Figure 1. Simplified model of stress and coping (Adapted from stress and coping models developed by Northouse LL, Caffey M, Deichelbohrer L, et al. The quality of life of African American women with breast cancer. Research in Nursing and Health 1999;22:449-460 and Wenzel L, Glanz K, Lerman C. Stress, coping and health behavior. In K Glanz, BK Rimer, FM Lewis, editors. Health Behavior and Health Education. San Francisco, CA: Jossey-Bass; 2003:210-239)



The Health eCommunities Study

The Health eCommunities study is a multi-method project, using both quantitative and qualitative methods, designed to assess the impact on cancer survivors and their caregivers of participating in mailing lists sponsored by ACOR. Health eCommunities represents a collaboration between the ACOR leadership and an interdisciplinary team from the University of North Carolina at Chapel Hill (UNC). An overview of the study and preliminary survey findings have been reported elsewhere [30].

Established in 1996, ACOR is a non-profit Web portal that offers users access to a rich array of information and support for cancer survivors, family members, friends, health care professionals, and researchers. The website also provides links to information about treatment options, clinical trials, and cancer-related books. As an organization, ACOR is a loose confederation of more than 150 publicly accessible mailing lists that range in size from very small (< 10 subscribers) to very large (> 2000 subscribers). The mailing lists are run by dedicated volunteer teams of listowners, most of whom are cancer survivors or surviving spouses or friends of deceased ACOR members. ACOR capitalizes on listowners' expertise by offering listowners their own mailing list in which to discuss shared problems in list management and facilitation. In addition to running their own lists, listowners may serve on one or more other ACOR listowner teams.

This paper focuses on the qualitative component of the study, including content and thematic analyses of email message texts exchanged by ACOR members (survivors and caregivers) in 10 lists. Recent technological innovations in automatic text analysis now make large-scale studies of support lists possible [37]. Our project is unique among the body of qualitative studies of mailing list–based email correspondence (most using manual coding) because it represents one of the largest samples of message texts of ESGs for related illnesses (eg, different types of cancer).

Because relatively little is known about ESG functioning, we began with two global questions: (1) What are the major concerns of ACOR members? and (2) What kinds of support do these groups offer? However, an additional focus of this paper reflects our unexpected finding that list members were more likely to send messages offering support than they were to request support. This finding was unexpected because previous studies of ESGs reported that most Internet users go online seeking both information and support; therefore, we expected that ACOR members would request support at least as often as they offered support. Further, list members who had the most direct need for support, the cancer survivors, were much more likely to offer support more often than listmates who were caregivers. For these reasons, this paper is devoted to an examination of cancer survivors' characteristics across the 10 ACOR lists, their patterns of supportive behavior, and the ways they used their collective Internet connections to promote their own and others' ability to cope.

Methods

Data Collection

Selection of Participating Mailing Lists

Although ACOR permitted our research team access to ACOR as an online health community, individual listowners made the ultimate decision whether their respective lists would participate. We used three criteria to select ACOR lists. First, they had to have listowners with sufficient experience and confidence in their roles as listowners to be good research collaborators. Using his knowledge of the listowners, Frydman identified likely volunteers, encouraged their participation, and worked with the UNC team to gain listowners' consent. Second, lists were

selected to represent a range of cancer diagnoses and likely prognoses. Finally, because the project's survey component would focus on new members, lists were selected based on their history of accruing new members. Although we knew that caregivers were active participants on these lists, we limited heterogeneity of the sample at the group level by excluding lists organized exclusively for cancer caregivers.

Sampling List Messages

For the qualitative study, we used individual email messages as the units of analysis. We sought to obtain enough messages to tap into the diversity of members' views, to capture differences in message content between very active and less active list participants, and to allow for stratification of survivor and caregiver data. Ultimately, we drew our data from the set of messages sent to the 10 lists over a five-month period between November 2003 and March 2004.

Mailing lists appear to have a life cycle of growth and decline, with the number of active members declining as the number of subscribers increases [38]. We did not have full baseline data on subscriber activity levels for the study lists in the months before the study began. One member of the team had been given permission to access the colon cancer archives for a previous study. We conducted pilot analyses of the colon cancer list participation data from November 2002 to March 2003 to estimate how many messages would be in the overall sample. Based on colon list data, we used a conservative estimate of one message per month per subscriber to define active membership.

Currently, there is no consensus among researchers about appropriate strategies for sampling the behavior of list participants. We chose to systematically sample messages sorted chronologically, assuming this procedure would enable us to tap messages from most active members and dominant discussion threads. Based on our pilot studies of the colon cancer list participation rates, we estimated that approximately 25% of subscribers would send a minimum of one message per month. Based on 9881 subscribers in the 10 lists in November 2003, we estimated that over 12000 messages would be sent during the five-month data collection period. Limited resources precluded analyzing all these messages; thus, we systematically sampled 9% of all archived messages for each group, anticipating a sample of 1112 messages.

ACOR staff used a multi-stage procedure to extract messages from the portal's archives. After sorting each month's messages chronologically, they used an automated system to extract every eleventh message sent by members (listowner messages were extracted in a separate step). Each month's message texts were compiled into a single email digest message and emailed to the UNC research team for review and analysis. Our data collection strategy was approved by the Institutional Review Board of the UNC School of Public Health.

Protecting Participants' Rights

By sampling messages from the ACOR archives instead of ongoing discussions on ACOR lists, we tried to avoid the possibility that participants would be deterred from posting to their respective lists, inadvertently disrupting their access to on-list support [39]

All listowners of the participating lists gave their permission to access their lists' archives. Prior to the start of the Health eCommunities project, they posted messages to their lists notifying subscribers about the project and directing them to the project's Web page for Frequently Asked Questions. Members were reminded that ACOR archives were accessible to the public.

Members were also told how sender anonymity would be protected. All message texts were de-identified by ACOR staff before the monthly message digests were transmitted to the research team. An automated process assigned each ACOR subscriber a unique ID number and removed subscribers' names and email addresses from message headers, message texts, and signature lines sent to the UNC team. Although we only coded original messages, for message texts that were replies to other messages, we retained the original message texts in the data set so we could correctly interpret the replies. Listmembers were offered an "opt out option." Those who wanted to have their messages removed from the data set could notify ACOR. In turn, ACOR staff would notify the UNC team about the list and participant ID numbers that would be affected.

Analytic Approach

As in most email communication, message texts in ACOR lists were informal and loosely organized [37], making analyses challenging. Messages also varied in length. Thematically, they were pastiches that could contain technical information about treatments, side effects, clinical trials, empathic comments, requests for information, or meta-comments about group processes.

Content Analyses

We conducted content analyses on each message to determine the relationship of the message sender to the cancer survivor (eg, survivor, caregiver, trusted other, or health care provider) and how often each sender corresponded with his or her respective list. A significant proportion of messages lacked sufficient information to deduce the sender's role. Subsequently, we reduced the number of messages with unknown senders by compiling all messages with the same sender ID numbers. In many cases, we could find one or more messages within those compiled subsets that contained enough information to recode all messages with identical ID numbers. During this process, we also discovered a number of cases where caregivers and survivors were sending from the same email address, as occurs when multiple family members share a common email address.

Frequencies of topics we report may not represent equivalent amounts of text. The unit of analysis was an entire message. We used Atlas.ti [40], a qualitative data analysis program, that allowed us to code each message text in many different ways. Thus, in a given message, a code could apply to a single sentence or many sentences depending on how extensively the sender addressed a topic.

For thematic analyses, we used both theory-based [41] and grounded theory [42] approaches. For theory-based analyses,

we initially developed global codes based on the sensitizing concepts of the project's simplified Stress and Coping Model (see Figure 1). Definitions for these coding categories were developed from items used in the project's online surveys. Following principles of grounded theory [43], we immersed ourselves in the texts, refined our code definitions, and developed new codes as new themes emerged. In general, global codes developed from the model captured the major domains of message content. However, over time, it became clear that we needed subcodes to categorize emerging components of the most frequently used codes.

We searched for overarching themes of support that reflected concerns of most groups. To be selected as an overarching theme on survivor supportive communication, a given topic had to be discussed in at least three messages in a majority of the 10 lists. To be included as a subtheme, a topic had to be discussed in at least three groups.

Maintaining Analytical Rigor

After testing global codes on pilot data, email digests from the lists were assigned to one of two coders (AM and EJL); each coder analyzed data from five of the lists. Following procedures of constant comparison, they met regularly to discuss and resolve coding issues. To determine consistency of codes across mailing lists, we conducted tests of interrater reliability on a 10% sample of each list. Reliability was assessed using a variant of the code checking strategy recommended by Miles and Huberman [44]. For each code, the percentage of agreement was calculated using as the baseline all the instances of text where at least one rater applied that code.

Results

Interrater Reliability

Miles and Huberman [44] recommend minimum interrater reliability levels of .70 for each code. Using this standard, our reliability levels were in the acceptable range. For the eight major codes, average reliability was .84. For the 44 subcodes, reliability was only slightly lower at .83.

Sample Characteristics

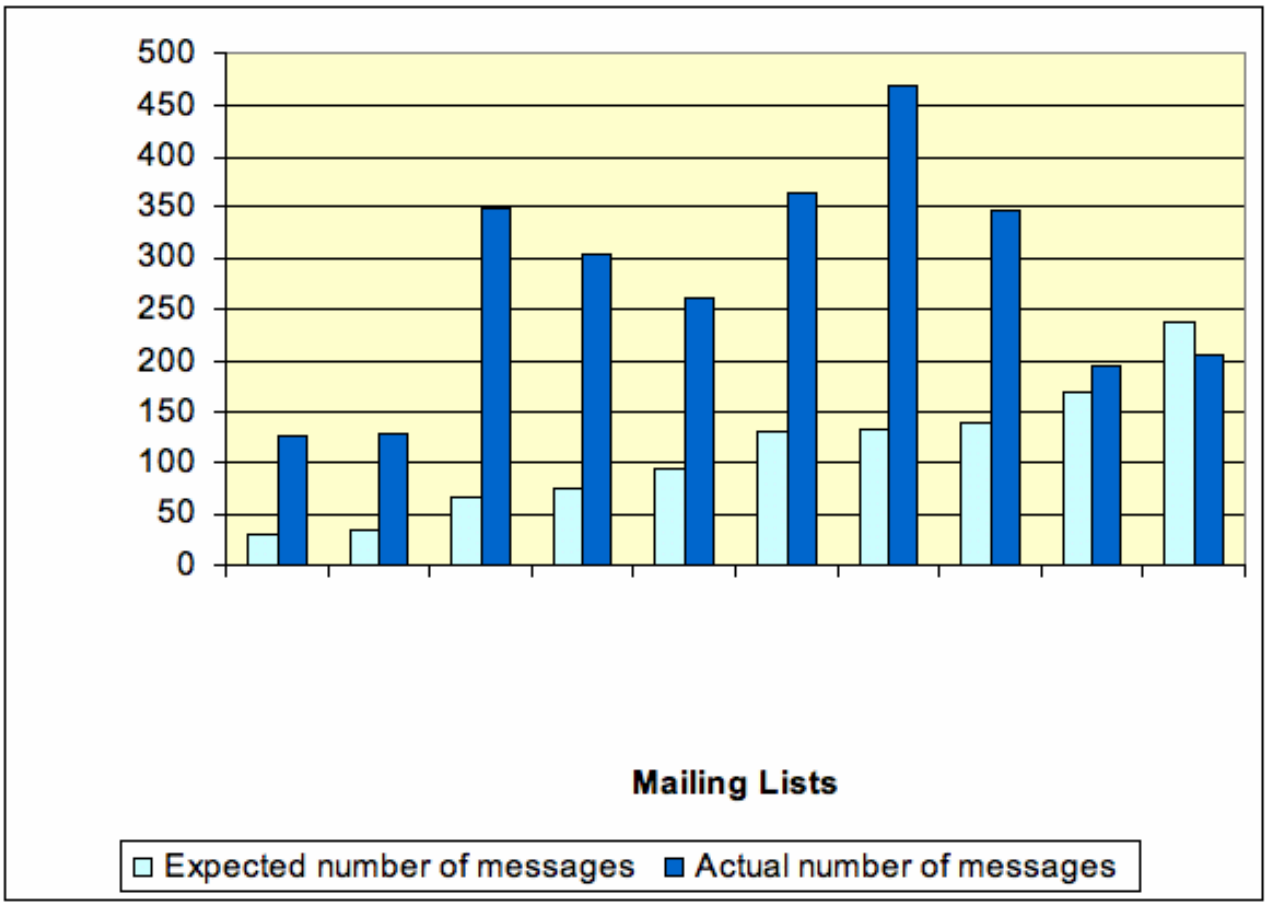
Sample size in online mailing list communication can be characterized along two dimensions: number of messages posted and the unduplicated number of message senders.

Number of Messages

Participants in the sample sent nearly 50% more messages (n = 2755) than our pilot estimates of 1112 messages (Figure 2). However, these overages were not evenly distributed across lists. The ovarian cancer list sample was more than 50% larger than expected (157%), and the leiomyosarcoma (LM sarcoma) list sample was over four times larger (427%) than expected; the chronic lymphocytic leukemia (CLL) list was the only one in which the number of messages was slightly lower than expected (−13%). There was no statistical relationship between average rates of participation and five-year survival rates as defined by Surveillance, Epidemiology, and End Results (SEER) data [45].

Figure 2. Participant activity levels: expected and actual number of messages

Figure 2. Participant activity levels: expected and actual number of messages



Number of Unduplicated Senders

Although subscribers can benefit from reading messages but not sending them (also known as "lurking"), in this study, we defined participation in terms of posted messages. When we started collecting data, the 10 lists had 10153 subscribers (Table 1). The number of individual senders varied by list from 63 (non-small cell lung cancer [lung NSCLC] and long-term survivors [L-T survivors]) to 164 (ovarian cancer). Smaller lists tended to have a greater proportion of active members.

Table 1. Comparison of mailing list size and number of senders in sample

Mailing List	Number of Members at Start-Up	Number of Senders in Sample (%)
L-T survivors	277	63 (23)
Lung NSCLC	310	63 (20)
LM sarcoma	587	95 (16)
Colon	670	115 (17)
Kidney	846	137 (16)
Esophageal	1160	127 (11)
Myeloma	1191	118 (10)
Ovarian	1503	164 (11)
Prostate	1503	108 (7)
Chronic lymphocytic leukemia (CLL)	2106	135 (6)

Participant Characteristics

Table 2 summarizes ACOR list participants' characteristics in terms of their relationship to cancer survivors. There were statistically significant differences in the composition of the lists in terms of the participation rates of types of participants (χ^2_{18} = 389.6; N = 2967; P <.001), with survivors posting more often than other types of list members. In the typical group,

messages from identifiable survivors made up slightly more than half (median 57%) of all correspondence, ranging from 38% (esophageal cancer) to 84% (ovarian cancer). Messages from identifiable caregivers made up slightly more than a quarter of the correspondence (median 27%), ranging from 5% (ovarian) to 48% (esophageal). Messages from other identifiable types of participants (friends and health care providers) were rare. Messages from participants who could not be identified typically made up slightly more than 13% (median) of the correspondence; this ranged from 8% (lung NSCLC) to 20% (CLL) of all messages sent.

Table 2. Mailing list characteristics: participation rates of different sender types and exchange of support

	Mailing List*										χ^2 (df) †	*P*
	CLL	Colon	Esophageal	Kidney	LM Sarcoma	L-T Survivors	Lung NSCLC	Myeloma	Ovarian	Prostate		
Participation Rates											389.6 (18)	<.001
Patient/survivor	126 (61)	152 (49)	140 (38)	124 (47)	229 (66)	99 (75)	60 (45)	248 (53)	449 (84)	126 (63)		
Caregiver	32 (16)	120 (39)	176 (48)	92 (35)	66 (19)	17 (13)	60 (45)	163 (35)	25 (5)	28 (14)		
Other	47 (23)	38 (12)	54 (14)	46 (18)	54 (15)	16 (12)	13 (10)	59 (12)	62 (11)	46 (23)		
Overall Support											66.17 (9)	<.001
Explicit requests	35 (39)	17 (17)	70 (50)	22 (25)	39 (28)	6 (12)	11 (25)	37 (24)	40 (18)	21 (26)		
Explicit offers	54 (61)	84 (83)	71 (50)	67 (75)	98 (72)	45 (88)	33 (75)	119 (76)	185 (82)	59 (74)		
Type of Support											33.14 (9)	<.001
Emotional	24 (26)	23 (22)	25 (30)	18 (19)	45 (30)	15 (29)	5 (11)	22 (13)	63 (29)	9 (10)		
Informational	69 (74)	82 (78)	58 (70)	78 (81)	107 (70)	37 (71)	40 (89)	141 (87)	153 (71)	78 (90)		
Emotional Support											8.88×10^{-8}	.105
Explicit offers	19 (79)	22 (96)	24 (96)	17 (94)	35 (78)	15 (100)	4 (80)	20 (91)	58 (92)	9 (100)		
Explicit requests	5 (21)	1 (4)	1 (4)	1 (6)	10 (22)	0 (0)	1 (20)	2 (9)	5 (8)	0 (0)		
Informational Support											30.10 (9)	<.001
Explicit offers	38 (55)	65 (79)	53 (91)	56 (72)	71 (66)	31 (84)	29 (73)	102 (72)	120 (78)	56 (72)		
Explicit requests	31 (45)	17 (21)	5 (9)	22 (28)	36 (34)	6 (16)	11 (27)	39 (28)	33 (22)	22 (28)		

*Numbers are reported as raw counts with percentages in parentheses. Percentages are of total messages for participation rates, total messages coded as containing support for overall support, and total messages containing emotional support and informational support for those two categories.

†Chi square values are Cochran-Mantel-Haenszel general association statistics with degrees of freedom in parentheses, except for Emotional Support, which uses Fisher exact test and reports table probability rather than a chi square value.

Mailing Lists as Sources of Social Support

Table 2 further shows the absolute numbers and percentages of messages containing explicit requests for and offers of support across all lists. We also found a substantial number of instances in which participants posted messages that could be construed as *implicitly* offering support although they were not specifically responses to others messages (range: lung NSCLC 28% to ovarian 53%). Such messages included descriptions of cancer-related events without comment or requests for help. Some senders wrote narratives about their diagnosis or staging, their experiences with different treatments, and treatment side effects. Other members posted URLs to link members with information about cancer treatments, clinical trials, or new research findings.

Some messages used automated message signature files that contained thumbnail histories of survivorship experiences of senders or survivor relatives (Textbox 1). Both survivors and caregivers used this signature line feature, and they included updateable, telegraphic lists of senders' (or their relatives') date of diagnosis, cancer treatments and outcomes, and current health status. Although some other support lists require members to include case history information in their signature lines as bona fides of members of their lists [A Jones, personal communication, March 1, 2005], ACOR lists do not require this. Of 1751 messages sent by identifiable survivors, 267 (16%) had case history signature lines.

Previous research on asynchronous Internet-mediated communication found that correspondents rapidly develop an idealized sense of intimacy [46]. We found the intimacy in our sample of ACOR lists tended to be focused on the pragmatic details of life with cancer. Furthermore, we found surprisingly few disclosures of events in members' lives that were unrelated to cancer survivorship.

Textbox 1. Four examples of survivors' signature line texts (in order, from briefest to more detailed)

№	Sample Signature Line Text
1	Rectal Duke B2; Cancerteer [sic] since 2000
2	[city, state]. (A 3-1/2 yr Fighting Four)
3	dx [date] from a Pap test, TAH/BSO, papillary serous adenocarcinoma, primary peritoneal, IIIa grade 3, 7 rounds taxol/carbo. Recurrence #I dx 9/99 from a Pap test, TAH/BSO, papillary serous adenocarcinoma, primary peritoneal, IIIa grade 3, 7 rounds taxol/carbo. Recurrence #I [date], 8 taxol/carbo. Recurrence #II [date] biopsy says estrogen + tamoxifen, 8 taxol/carbo.
4	[date] - X-rays showed spot on left lung [date] - Surgery tumor & lower lobe-left lung removed; not clear margins; 5-6 nodules like grapes 15 x 10.5 x 8.0 cm mitotic rate was 4-5/50 hpf [date] - Cat scan chest & abdomen showed tumor in or near left lung [date] - Cat scan of Head and Pelvis clear [date] - Surgery 1 tumor removed from lining of ribs; not clear margins; surgeon recommends radiology; Size: 2.8 cm x 2.2 cm x 1 cm; mitotic rate. increased to 2-3/10 hpf; & more cellular specimen. [date] - Sarcoma Clinic dx: Smooth muscle tumor of uncertain malignant potential/grade 1 leiomyosarcoma (from oncologist report to my surgeon & GP). No stage reported: when pressed Medical Onc. Said maybe 2 or 3. [date] - No adjuvant therapy; 6 mths NED; Rad Onc says Low Grade LMS needs close follow up. My surgeon will attend to follow up Scans and Xrays from now on. [date] - xray in Sept clear 9 months NED [date] - CatScan to be done in Dec/03

Information and Advice Themes

Because our surveys of new subscribers found that cancer survivors join ACOR lists mainly because they are looking for information , it was not surprising that this theme was dominant across lists in our sample (see Table 2). The most common kind of support offered by survivors was information and advice based on their experience. We identified four major themes associated with survivors' offers of information and advice: (1) specific treatments, (2) communicating with health care providers to find the best treatment, (3) problem management strategies, and (4) coping with cancer recurrence. Each of these themes was subdivided into several related components. Table 3 lists the major themes, associated subthemes, and the total number of lists in which survivors sent supportive messages addressing each theme.

Our sample of messages included relatively little discussion about the experience of being diagnosed. Although survivors' messages with case history signature lines often included the diagnosis date, slightly less than 10% of messages (79 of 811) from survivors mentioned their diagnosis experience.

Table 3. Information and advice themes

Theme Number	Themes and Subthemes*	Number of Lists †
1.1	**Specific treatments**	**10**
	Case histories	10
	Treatment types	9
	Factors to consider in making treatment decisions	9
1.2	**Communicating with health care providers to obtain good care**	**9**
	Communication factors that could affect quality of care	7
	Strategies for obtaining good cancer care	5
1.3	**Problem management strategies**	**8**
	Using the Internet for due diligence	8
	Using the Internet to get good cancer care and social support	4
1.4	**Coping with cancer recurrence**	**6**
	Coping with risk of recurrence	5
	Living with relapses	3

*To be selected as a major theme (in bold), a given topic had to be discussed in at least three messages in a majority of the 10 lists. To be included as a subtheme, a topic had to be discussed in at least three groups.

†Numbers of discussion subthemes in lists do not necessarily equal the total number of mailing lists in the major theme. These counts represent the union of the set of *all* mailing lists in which messages addressed some aspect of the major theme, while subtheme occurrences may only have been found in a subset of the mailing lists.

Theme 1.1: Specific Treatments

Most new subscribers reported that one of their main reasons for joining ACOR lists was to seek information and guidance about specific treatments [30]. Theme 1.1, Specific Treatments, dominated discussion in all 10 lists. There were three subthemes: case histories, information about types of treatment, and factors to consider when making treatment decisions.

In all lists, members often discussed their case histories. These messages contained accounts rich in information that other members could use to estimate the likelihood of long-term survival. In nine lists, survivors offered information about treatments and factors to consider when making treatment decisions. In discussions of treatment types, senders provided information to help listmates understand the bewildering array of options; of these, chemotherapy was mentioned most frequently. In about a third of the messages, members discussed two or more kinds of treatment options and frequently discussed how they were used in combination.

Survivors' messages often offered help to listmates in thinking through the factors to consider when making treatment decisions and typically urged the newly diagnosed to practice "due diligence" by becoming informed consumers of cancer care. Due diligence included seeking information on treatment response rates, probability of survival or cure, likelihood of recurrences, long-term survival rates, and cancer stages. To be informed consumers, newly diagnosed members were advised to seek information on the types and severity of side effects and after effects of treatment and the likelihood of cure associated with the specific treatments being proposed and relative to other available treatment options. In some cases, senders provided information they thought listmates needed, describing their experiences with treatments, side effects, and strategies for minimizing discomfort. Senders also provided assistance for those further along in their disease progression and treatment cycles. Here, discussions addressed such diverse issues as how risk of recurrence is assessed, how to use test results to decide on whether to use standard or aggressive treatments, potential benefits of adjuvant treatments, and sequencing of treatments for different stages of cancer (eg, reserving certain chemotherapy regimens in the event the cancer became inoperable).

Theme 1.2: Communicating With Health Care Providers to Obtain Good Care

In nine lists, survivors offered information and advice on issues related to the role of good communication in obtaining optimal cancer care. This theme included two subthemes: the importance of communicating with health care providers and recommendations about how to obtain good cancer care.

In seven lists, survivors discussed communication factors that could affect quality of care. They advised listmates to take initiative and raise issues with their doctors to understand the relationship between disease progression and treatment options, to be prepared for treatment side effects, and to understand how to maximize quality of life. Senders provided examples to distinguish between good and bad communication with health care providers, and they described how their relationship with their provider had benefited through good communication.

Discussions about problematic communications with health care providers occurred in half the lists. Members shared examples of poor communication and bad advice received from health professionals. They offered suggestions for critically assessing the information and advice they received and encouraged listmates to use lists to clarify and validate

professional opinions. Some senders posted messages expressing concern about the information or recommendations other members reported from their doctors, urging them to seek second opinions. They also offered suggestions for coping with adverse consequences of inappropriate treatment decisions.

In five lists, members offered advice about strategies to obtain good cancer care and identified treatment centers and doctors who provided good treatment. Senders offered advice on building a good medical team, how to work with team members, and how to cope with the complexity of information from different team members.

Theme 1.3: Problem Management Strategies

Survivors in eight lists shared information and advice about different problem management strategies. There were two subthemes: using the Internet for due diligence and using the Internet to get good cancer care.

List members had already enacted one problem management strategy by joining an ACOR list and encouraged each other take further action against the cancer threat. Survivors' messages provided information and advice about effectively using Internet resources to obtain and manage cancer care. Senders clearly viewed the Internet as an essential tool for cancer-related problem management.

Information messages from survivors supported listmates in using the Internet for due diligence by providing specific content of interest to individuals with a similar cancer diagnosis. Senders posted hyperlinks to websites they thought salient for listmates. The most frequently mentioned websites included those that provided definitions of diagnostic terms, descriptions of treatment protocols, and information on treatment side effects. In some cases, websites were sponsored by the National Cancer Institute (NCI) or other government agencies. Members also frequently recommended other nongovernmental sites that offered information about clinical trials or cancer-related research projects, links to published reports of research findings, and facilities offering specific types of cancer treatment. In four lists, survivors posted messages in which they recommended online publications or books they believed were reliable and helpful sources of information. When members recommended books, they typically included links to online retail sites offering the book.

The second subtheme, using the Internet to get good care and social support, was found in four lists. Typically, these survivor messages shared information about how to use the Internet to get the best possible care for particular cancers. Survivors also directed listmates to websites that contained information about providers from whom senders had received good care. In addition, survivors helped listmates find sources of social support, including other cancer support lists. At the same time, senders reinforced listmates' participation in ACOR mailing lists by noting how to use listmates as resources. Senders often said they valued ACOR mailing lists because they provided access to more experienced members who could help them interpret test results and correct misinterpretations about published cancer research studies. In addition, senders noted how ACOR listmates provided practical information about the daily struggles of coping with cancer, coping at different stages, and the realities of long-term survivorship, all of which was a valuable supplement to the advice received during all too brief contacts with health care providers.

Theme 1.4: Cancer Recurrence

If survivors are fortunate, their treatments result in cures or very long periods of remission. However, being declared disease free does not mean these individuals can return to their precancer lives, so both fear of relapse and coping with relapse were important concerns. The major theme of cancer recurrence was discussed in six lists and included two subthemes: managing risk of recurrence and coping with relapses.

In five lists, survivors discussed problem-focused coping with the ambiguity of being found NED (no evidence of disease) or NERD (or no evidence of recurring disease). They admitted that they remained anxious about the possibility of recurrence, and reminded their listmates that, even after being declared disease free, they must remain vigilant for early signs and symptoms of recurrence. They also advised using due diligence at this stage by being proactive and staying informed about new treatment options, probability of cure and recurrence, and treatment side effects and after effects. They encouraged listmates to take steps to minimize the risk of recurrences. For example, senders recommended lifestyle changes, such as dietary changes, to reduce the likelihood of recurrence of colon polyps. In some lists, senders wrote about the importance of regular self-monitoring, noting that occurrence of a primary cancer increases the risk of subsequent ones, sometimes in different forms. They shared information about procedures for monitoring and discussed how to cope with the uncertainty of fluctuating test results and how they could use test findings to decide whether to seek treatment to prevent a recurrence. In one list, a member discussed the difficulties inherent in using statistics to assess an individual's prognosis and the importance of regular monitoring:

> *Since [cancer type] patients come in all shapes and sizes, some with very aggressive disease on diagnosis, others smoldering for many years; it's impossible to have a level playing field.... My internist FIRST noticed a slightly elevated protein level when I became his patient…. Prior to that, no tests were ever taken that would break down this figure. It wasn't until [date removed] that this doc felt the protein level needed to be checked by a hem/onc, who decided that I was "smoldering." I finally started treatment in [date removed] and had a transplant in [date removed]...one of the first combined stem cell and bone marrow transplants. Though I have relapsed twice, I'm still kicking.*

In three lists, survivors reported how they coped with relapses. Although all felt emotionally devastated when informed about their recurrence, senders agreed the only way to cope was to keep in mind that they had to deal as well as possible, which included getting all the information they could. One member offered an example of how staying informed about treatment options enabled her to help her children adjust to the reality and implications of her relapse. Other survivors, such as the

participant quoted earlier, provided information to sustain hope when confronted with relapses, reminding listmates that—despite multiple relapses—they were "still kicking."

Emotional Support Themes

Defining Emotional Support

Compared to the detailed technical discussions of cancer treatments, the emotional supportive text of survivors' messages was often relatively brief and typically occurred at the end of messages. To categorize the emotional aspects of support survivors offered to others, we adapted Cutrona and Suhr's [47] typology of emotional support (Tables 4 and 5).

Table 4. Categories of emotional support (adapted from [47])

Type of Emotional Support	Description
Emotional coping strategies	Suggestions for managing painful emotions (eg, anxiety, anger, fear, and sadness)
Empathy	Acknowledging identification with others' emotional reactions and feelings including painful and pleasurable feelings; validating the appropriateness of another's reactions to stressful circumstances
Encouragement	Support in persistence in facing challenges; expressing hope that situations will improve
Prayers	Offering spiritual support through prayers or blessings to others in distress
Esteem support	Appreciation for the value of an individual and his or her accomplishments

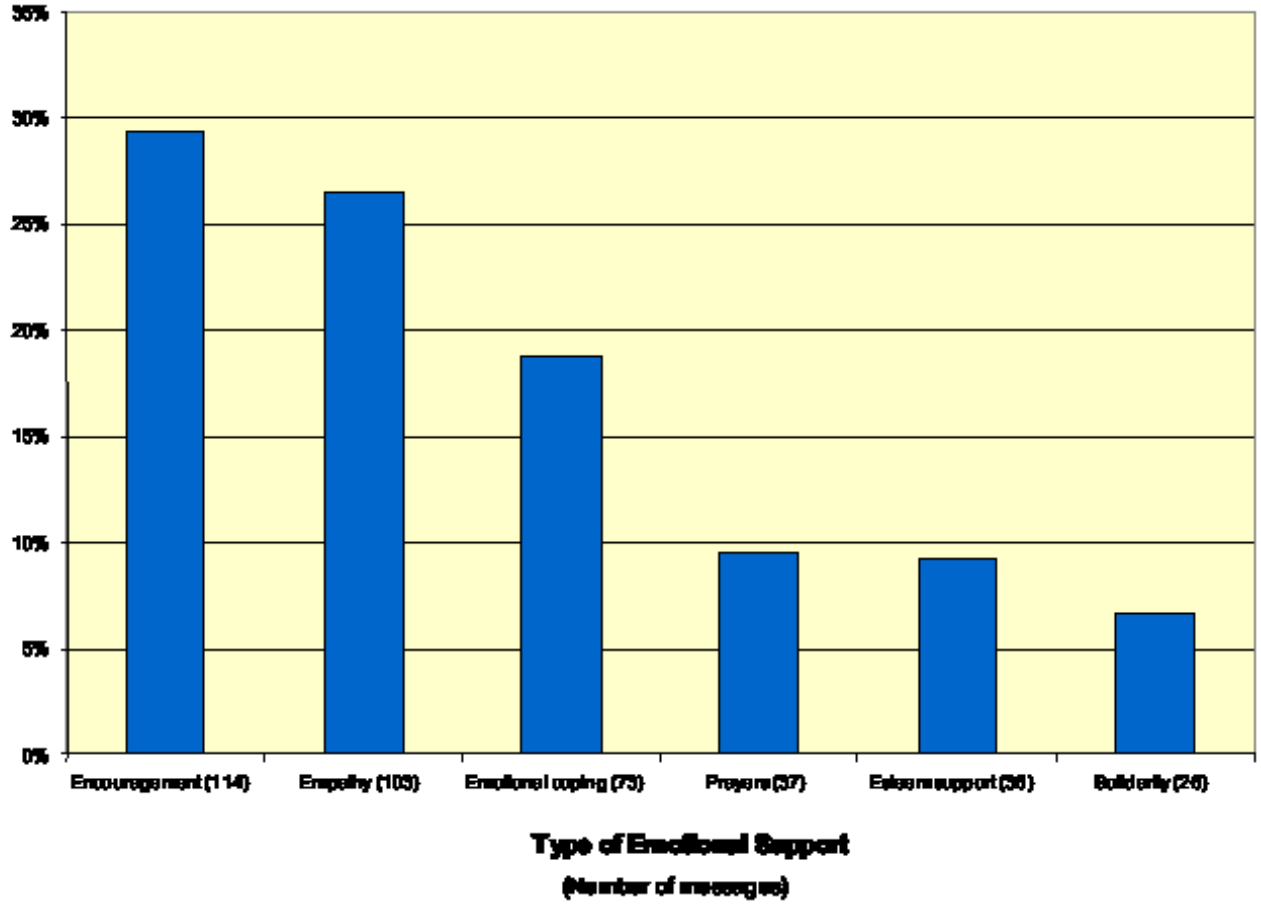
Table 5. Emotional support themes

Theme Number	Theme	Number of Lists
2.1	Encouragement	10
2.2	Empathy / emotional validation	10
2.3	Emotional coping strategies	10
2.4	Esteem support	9
2.5	Prayers	7
2.6	Solidarity	7

Rates of Emotionally Supportive Communication Across Groups

As Table 2 shows, the 10 lists can be roughly divided into two groups along the level of emotionally supportive communication. Seven groups (L-T survivors, CLL, LM sarcoma, kidney, colon, and ovarian) were more supportive, with 19% to 30% of messages containing emotionally supportive comments. Emotionally supportive messages in the three less supportive lists (myeloma, lung NSCLC, and prostate) ranged from 10% to 13% of all messages. In addition, types of support varied across lists. Encouraging and empathic comments, as well as suggestions for emotional coping strategies occurred, to some degree, in all groups. In addition to encouragement and coping strategies, members in a majority of the groups offered several other kinds of emotional support (Figure 3). However, compared to encouragement and emotional coping strategies, the other types of emotional support occurred much less frequently.

Figure 3. Types of emotional support offered by survivors (N = 389 messages)



Theme 2.1: Encouragement

As befits support groups, the most common type of emotional support was encouragement; about a quarter of messages included some encouraging notes. Frequently, these were simple exhortations at the end of messages, such as "Hang in there" or, for those embarking on a course of treatment, "Hope everything goes well for you!"

Others offered warm wishes for continued good recovery and "continued shrinkage" of tumors. Many messages encouraged the intended recipients to take a more activist attitude in the face of difficult circumstances. In all groups, members encouraged each other to persevere and to persist in staying vigilant for recurrences, seeking appropriate treatments, and overcoming obstacles to getting Social Security support benefits. A fourth type of encouragement served to sustain hope that cancer is treatable, that the intended recipients are "in good hands" at their respective treatment facilities, and that feeling better and continued good quality of life are achievable goals.

Theme 2.2: Empathy and Emotional Validation

Empathy is the ability to understand and identify with another person's feelings in both happy and sad occasions and to respond compassionately to another's distress [48]. We categorized empathic statements as those in which survivors wrote messages such as "The things you write are the same we've been through" and "We know how it feels to fight this beast." We included comments that validated other members' emotional reactions to the experiences as cancer survivors, such as "Like you, I am always scared before getting test reports" and "You can be pissed off and let off steam." Survivors also expressed sympathy for each others' disappointments, most often related to reports of recurrences or metastases. Happily, survivors also wrote when they had good news, such as good test results, announcements that they had survived yet another year, weddings, births, or other positive events. Listmates responded to these messages with expressions of pleasure and sincere congratulations, often mentioning how important hearing good news was to sustaining their hope for the future.

The most solemn types of emotional support were expressions of sympathy and condolences offered to caregivers and their families when loved ones died. These occurred primarily in three lists (LM sarcoma, L-T survivors, and ovarian). In instances in which the deceased had been an active member of the list, listmates sometimes responded with moving tributes: "[name deleted] was such an inspiration to us, always sending articulate and intelligent posts. We learned so much from our association with [name deleted] and the knowledge he shared."

Theme 2.3: Emotional Coping Strategies

ACOR list members provided models of successful emotional coping by offering descriptions of attitudes and cognitions that helped them in their journey. Many messages encouraged self-acceptance, emphasizing that each person's experience is unique and there are "no right or wrong ways to do it." Similar advice on self-acceptance was offered for coping with treatment side effects and after effects: "Accept the new you" and "I

enjoyed having my straight hair come back curly after treatment."

Members frequently wrote about how their life view was transformed by the cancer experience, moving from dread to deep appreciation for the preciousness of life and the people they loved. One member described how keeping a journal of her cancer journey helped her find new meaning in life: "Strange that we have to suffer to find ourselves." Members consistently stressed the importance of maintaining a positive attitude, and many wrote about the importance of experiencing life more fully. Members reminded each other to not only plan for the future, but also live life for today and not defer pleasure, such as long-planned vacations.

Members also described positive coping by using social comparisons. They could tolerate treatment side effects if it kept "the beast at bay," bolstered by the prospect of feeling better. Listmates offered examples of people who had lived long lives and accomplished much even when living with cancer.

Members also offered practical advice. One suggested that listening to music could reduce feelings of claustrophobia while being scanned. Others recommended managing anxiety by recalling images of favorite peaceful places or seeing a psychiatrist to manage guilt and the anxiety of "living in limbo after chemo."

Theme 2.4: Esteem Support

We defined esteem support as verbal expressions that specifically built a sense of self-worth, value, and competence. Survivors nurtured their listmates' self-esteem by explicitly acknowledging how frequently their messages served as sources of expert information, strength, and inspiration, as well as models of courage and coping. Various members commented how messages had helped them cope at different stages of their illnesses: when newly diagnosed, when confronting a new course of treatment, and at the terminal stage (for example, "How you inspire us with your courage and wisdom as you face what we all must."). In addition, members expressed appreciation for those who continued to participate in the lists and provide support for others even after they had been declared NED (no evidence of disease). Survivors also commented appreciatively about listmates' individual characteristics, such as courage, assertiveness, humor, and a positive attitude, thus reinforcing strengths that would help them cope with their cancer. Not surprisingly, members also valued good writing and members who could "verbalize what so many of us feel."

Theme 2.5: Prayers

Survivors who offered help were relatively secular in their support. Across all groups, prayers and blessings were relatively rare (Table 5), and we did not find any in three of the lists (kidney, lung NSCLC, and prostate). In contrast, almost half of survivor messages sent in the ovarian list contained a reference to prayer. The type of spiritual support offered in these messages tended to be brief, conventional (although certainly sincere), and included phrases such as "God bless you" or "You and your family are in my thoughts and prayers" as part of the message closing. In the section on listowner facilitation roles below, we discuss factors that may have deterred members from expressing more religiosity.

Theme 2.6: Solidarity

We found the final type of emotional support, solidarity, scattered across survivors' messages in seven of the lists. This type of support included general expressions of solidarity so members would not feel alone in their treatment tribulations. As with prayers, half the messages containing expressions of solidarity were sent by members of the ovarian list. Members wrote expressing interest in learning how others in the list were faring with treatment, but most often these expressions occurred as simple statements of support such as "I will be thinking of you" or "I wish the best for you."

Listowner Facilitation Roles

Across all mailing lists, discussions appeared to be very task focused. Neither survivors nor caregivers shared much information about their families or jobs or other aspects of life not directly affected by cancer. This raised the question of how members learn what is appropriate for discussion. We queried ACOR listowners by email about this issue, and they informed us that they worked both online and offline to facilitate discussions (Table 6). ACOR mailing lists function as a loose federation; listowner teams on each list have the authority and flexibility to shape their respective lists according to their shared personal visions and understanding of list needs. Listowners also have a closed-membership list of their own called Oncolist. Participation in Oncolist provides listowners with access to the knowledge and experiences of other listowners specific to facilitation of cancer-related mailing lists.

Although the content of each list varies substantially, ACOR provides a framework for some standardization. For example, all new subscribers to ACOR lists receive an automated welcome letter from Gilles Frydman, ACOR's founder. In addition, each list sends out an automated welcoming message that describes the group's intended audience as well as its purpose, but it generally does not specify topics that are unacceptable to the list community. In some instances, listowners have added brief, automated signature line reminders about list netiquette, such as this posting from the lung NSCLC list: "Support is our goal. Please, no: Commercial, religious, political, rude, or off-topic messages; no unapproved questionnaires."

ACOR listowners who participated in our study reported that they preferred not to actively direct discussion content because it could be disruptive and provoke flaming.

Table 6. Listowner facilitation roles

Role	Description
Modeling appropriate Behavior	We try to model appropriate behavior by offering praise and thanks off-list to those members who have been helpful to others. While we don't "coordinate" these efforts, we do copy each other on our private messages to members so that there is less chance of overlap/duplication. I think we also "model" appropriate list behavior with our own posts to the list (by, for example, staying on topic, trimming prior text, and the like). (ES, personal communication, November 10, 2005)
Keeping discussion focused on cancer-related topics	We also try to limit humor and cute stories as much as possible, because while some people like them, there are other venues for that and they take up a lot of space. Anyhow, they irritate all three listowners and it isn't a democracy ;-). (DB, personal communication, November 11, 2005)
Enforcing group norms	I do occasionally send out private messages to individuals that have posted inappropriate material…. Usual causes for personal messages are posts containing personal attacks of individuals, expressions of inflammatory political opinion, commercial content, etc. Haven't seen any controversial religious stuff in quite a while. I do feel that when the discussion establishes at a level of higher quality that it tends to sustain itself and little intervention is need on my part. If an eccentric individual begins posting inappropriate or provocative material, this tends to invite response, and dialog can deteriorate. (CC, personal communication, October 28, 2004)

List members often remind each other about rules, reducing the need for active interventions by listowners. Mailing lists that are part of a portal have an advantage when they encounter a member who persistently goes off topic, because they can easily redirect that member, encouraging him or her to join a list elsewhere on the portal where members' interests would be more convergent. For example, if a member frequently posts religious messages beyond conventional requests such as "Please pray for me," listowners may urge the member to join another ACOR list, such as the Cancer Survivors Christian Online Support Group, which specializes in Christianity and cancer. Similarly, members seeking support for use of alternative cancer treatments have the option of joining the CAM-ONC list, which provides a forum to discuss how to integrate complementary and alternative cancer care with conventional approaches.

Discussion

We found consistent patterns across the lists in participants' behaviors, focal topics of discussion, and in the types of support they offered each other. In this section, we focus on: (1) factors related to supportive discussion content across lists aimed at helping list members cope with their diseases and treatment effects, (2) methodological issues related to project sample size and participant characteristics, and (3) new automated methods to help researchers analyze the large quantities of data generated by online communities.

ACOR Mailing Lists as Sources of Support

Implicit Support

An important component of our investigation was to examine the functions of implicitly supportive messages and the role they play in overall support offered through lists. Preece argued that personal narratives and case histories are examples of empathic communication [49]. Unsolicited case histories can strengthen the group's solidarity and reduce individual members' sense of isolation by allowing other members to know how others are weathering their cancer-related challenges. Further, when members send messages that include URLs of resources, they remind the list of resources that are accessible via the Internet, making it easier for their listmates to engage in coping using problem management strategies.

Dominance of Offers Over Requests

We were surprised to find large and consistent differences in the proportion of messages offering support compared to those explicitly requesting support. However, based on the few studies that explicitly described list-based offers and requests for support, this behavior may be the norm rather than the exception [21,50,51]. Although researchers have presented their data in ways that make it impossible to make direct comparisons, there are conceptual similarities. In studies examining support-seeking behavior, the percentage of messages containing offers of informational support ranged, at the lower end, between 37% (problem drinkers) [21] and 40% (primary biliary cirrhosis) [50], and at the higher end, between 80% (colon cancer) [52] and 85% (diabetes) [51]. Our findings are consistent with those of earlier researchers who found offers of emotional support were less common, ranging from nearly 29% (problem drinkers and diabetes) [21,51] to 48% (colon cancer) [52]. Similarly, in these same studies, there were proportionately more requests for information, ranging from 15% (problem drinkers) [21] to 23% (colon cancer) [52]. Messages containing requests for emotional support were rare in all these studies, ranging from 2% to 3% of all messages. Populations and methodologies in these studies were so disparate as to make it difficult to draw strong conclusions about these differences.

We cannot determine the reason for these patterns. One possibility is that many ACOR list members do not have to ask for help because their offline resources for support are adequate. Within ACOR they can get answers on many topics or find role models for coping simply by reviewing list archives or lurking long enough to read exchanges of messages between others who are facing similar challenges [53]. Also, email communication is asynchronous and allows senders to remain invisible and anonymous. These restricted channel characteristics promote idealized views of senders [46]. Further, senders can choose their words and manage how they present themselves to the lists. Given the choice, it is likely that most people would choose to present themselves in a positive light. However, many of these people were under extreme stress, and that tends to reduce

the likelihood of positive framing. Finally, being able to offer help is a more empowered position than being a supplicant. List members who have more personal experience with cancer and have learned a lot about cancer from participating on the list can find rewarding roles as "elders," sharing both what they have learned and how they have coped [54].

Factors Affecting the Range of Discussion Topics

Information and Advice About Cancer Treatments

The combined findings from prior research on ways people use the Internet, the stated missions of mailing lists, as well as new ACOR subscribers' self-reports about what they were seeking by subscribing to the lists led us to expect that the dominant themes would focus on cancer-specific treatments. This hypothesis was borne out by our analyses. However, ACOR members presented this information in ways that suggest they were inventing their own form of evidence-based medicine. Members shared information and advice about clinical trials, their own experiences with side effects and after effects of specific treatments, and the practical points of managing different treatment [55]. With this information, list members felt better prepared to make informed choices, to become active members of their own care teams, and to negotiate with doctors and medical teams about which treatments to pursue. Similarly, by being better informed they could cope more effectively with the effects of treatment as well as with recurrence when treatment failed. Nevertheless, we did not have data by which to evaluate the quality of decisions participants made.

Typically, discussions about qualitative research findings remind readers that findings are based on convenience samples and are not generalizable to other situations or populations. However, we believe we can make stronger claims about the external validity of our conclusions. Our results are supported by the results of our quantitative Health eCommunities surveys of new ACOR subscribers, which found that they were most interested in informational support. Our results are also consistent with findings from several other studies that included analyses of list member communication across diverse health support lists whose discussions focused strongly on treatment-related issues [21,51,54,56-58].

Communicating With Health Care Providers and Finding the Best Treatment

People are unlikely to subscribe to illness-support lists unless they—or people they care about—have an illness. List members' messages were somewhat more likely to address issues about what to talk about with their doctors than about finding treatments. Many list members had already been diagnosed and were embarked on some kind of cancer treatment journey. Members reported examples of both good and bad relationships with their treatment providers. In general, discussions about finding treatment providers were more often about doctors and institutions offering the *best* treatment.

In the majority of lists, members focused on what they should discuss with their doctors. In five lists, some members expressed concerns about recommendations their listmates received from physicians and advocated seeking second opinions. In the literature, there have been reports on the reactions of providers whose patients bring information obtained from the Internet to the clinical encounter [59]. However, in our data, there was little reporting about list members bringing information and suggestions from mailing lists to their doctors or about how their doctors responded. It might be helpful to develop materials for listowners with suggestions about how to communicate optimally with physicians. Such information could be shared among lists.

Virtual Community Support for Active Coping

In our conceptual model of stress and coping (see Figure 1), list members' levels of self-efficacy and their information-seeking styles (illness appraisals) play a role in their coping strategies. For message analysis, we lacked baseline assessments of these attributes. We hypothesized that participation in the lists and social support provided by lists would impact coping efforts. Because we did not have pre- and post-measures of individual illness appraisals or coping styles for these quantitative analyses, we could not determine the extent to which participation in lists caused members to change their coping behaviors. However, these data show that when cancer-affected people join ACOR lists, they are consistently encouraged to engage in active coping behaviors (eg, information seeking, asking questions of health professionals). Members encouraged each other to maintain this "activated" stance at each stage of their treatment, regardless of whether their cancer was active or in remission. Moreover, list members were provided models of how they could use the Internet for due diligence in learning about their particular form of cancer, the most effective treatment options, and the institutions or providers from which they could obtain good care.

Emotional Support

List members offered each other encouragement, empathy, and shared emotional coping strategies. Given how frightening the prospect of cancer and cancer treatments can be, it is somewhat surprising that comparatively few of the messages contained explicit, emotionally supportive content (7% to 18% of messages). In addition, although messages were relatively restrained in emotionality, we found several kinds of emotionally supportive content in the messages, possibly increasing their impact.

Overall, survivors' emotionally supportive messages reinforced list values for active coping strategies. Survivors often couched their messages in optimistic, activist, or even militant terms (eg, "Never give up the fight!"). Moreover, messages offered esteem support for active members who modeled active coping. Survivors who offered support also recognized that there were situations in which active coping strategies were not appropriate. They acknowledged that coping strategies had to fit individual situations, modeling cognitive reframing strategies to cope with existential situations rife with uncertainty or many uncontrollable factors. ACOR members tried to provide help by suggesting coping strategies appropriately tailored to their listmates' situations.

Methodological Issues

Ethical Issues in Qualitative Research on ESGs

Among researchers studying online groups, one of the ongoing debates is whether mailing lists and other online groups are "public spaces or private rooms" [39,60]. This issue has major ethical implications for how qualitative research is conducted in these groups. Eysenbach and Till [60] have suggested that when some form of registration is required to gain access to the list, then most subscribers are likely to regard the group as a "private place." They also note, however, that members may view groups with a larger number of subscribers (> 100) as less private than those with fewer subscribers. Qualitative researchers who want to use email message texts from existing groups are confronted with the challenge of obtaining informed consent from members who did not join the lists to participate in research. While researchers may be able to obtain listowner permission to use messages, this strategy does not allow individual list members to give their consent. Once the data are analyzed, if verbatim quotes from members' messages are used in publication, then the researchers are obligated to reduce the risk of unwanted exposure by removing all identifying information from message texts and to obtain the senders' permission to use them.

In this study, we made systematic efforts to protect participants' rights at each stage of the research. We only recruited larger lists (> 200 subscribers). Although new subscribers to ACOR lists have to register, the ACOR website informs all newcomers that that ACOR archives are publicly accessible. As a result, participants in this study were more likely to view their lists and list archives as public rather than private spaces, and less likely to resent having their messages used in research.

With regard to consent, we had listowners' permission to use messages from their archives and we notified subscribers about the project's online fact sheet. The scale of the study prevented us from obtaining individual consent from all subscribers. In any case, members could request to have their messages removed from the analysis, but none did.

Although we did not promise to request permission to use quotes, the sheer scale of the project protected participants' privacy. The goal of the study was to identify commonalities in the provision of support across lists so the data were summarized extensively. Ultimately, we used very few direct quotes and most of those were only brief phrases. In the few instances in which we present more extended quotes, we removed information linking the quote to cancer types or specific lists.

Sample Size

Posts were highly variable in timing and frequency. It was an empirical question whether systematic sampling would capture this variability. As it turned out, we underestimated activity levels within lists (see Figure 2). On a gross level, comparing mean activity levels between the sample and full data sets, sample activity levels were higher in 7 of 10 lists. In general, these average differences ranged from one to two messages overall per month, with the one exception of the myeloma list, where the difference was approximately five messages. The ovarian list was the only list in which the sample averages were somewhat lower than those of the full list dataset.

We do not know the reasons for these differences between the sample and full data sets. One possible explanation is that, when more active members went online, they initiated more new messages *and* posted replies to other frequent senders during relatively short time periods. With our systematic sampling method of extracting messages from chronologically sorted sample messages, we may have tapped into a somewhat greater proportion of those active members' posts.

Differences in participation rates between groups may also have other implications for the way support is experienced by members in different lists. If the typical member only posts one or two messages a month, they may be less likely to use those messages to comment on or to reply to others' messages. A consequence of infrequent reading of list email is that individual senders are less likely to see replies made to their own postings. In turn, listmates whose response fail to elicit follow-up comments or answers to a posed question may come to regard nonresponsiveness as the norm for the larger mailing list community. As a result, members may experience the list as less responsive to individual needs.

Participant Characteristics

Group composition was another source of variability across lists. Similar to many other Internet-based organizations, ACOR does not collect background information on its subscribers. Thus, we did not know whether subscribers were survivors or cancer-affected others. We could only deduce their identity as survivors or caregivers from what they wrote in messages. We believe we were successful in doing so in most cases. The median number of messages for which we could identify senders' roles was 87%, but this proportion varied across lists. We were able to identify the most sender roles in the lung NSCLC list (92% of messages) and the fewest in the CLL list (80% of messages).

We found, as expected, that the majority of messages were from cancer survivors. However, there was also considerable variability in the proportion of messages sent by those whom we identified as family caregivers, ranging from a low of 5% in the ovarian list to nearly half (48%) in the esophageal list.

Lists with the highest proportion of caregiver messages were also the ones serving survivors with the most potentially deadly cancers (eg, esophageal). This finding is consistent with other research showing that Internet users who were support seekers were more likely to be caregivers than people suffering from health problems [6]. Possibly, these situations occurred when the survivors who originally subscribed to the list became too ill to post messages, so their caregivers took over. Also, as survivors get sicker, their caregivers may need more support and so post more frequently to their respective lists, thereby increasing the likelihood they would be included in our sample. If caregivers feel isolated due to demands of care giving, lists may serve an especially important role. As with other structural variations in ACOR lists, our explanations for these differences are speculative.

Although introductory information in list welcoming messages specifically mentions that health care providers are encouraged to participate, we found little evidence that they did. For the most part, those identified as health care providers were also cancer survivors. In a few cases, information in message texts or footers suggested that the health care provider sending the message had drawn on both personal and professional experience to become a cancer patient advocate or was involved in producing educational materials for patients.

Prior research indicates that many health professionals are concerned that patient-provider relationships will be affected when survivors bring information found on the Internet to discuss during appointments [59,61]. If health professionals participate in health support mailing lists as participant observers, they may become better informed about how lists function and may be more effective in helping survivors maximize the benefits of list participation. Surveys of physicians [61] and new ACOR members [30] showed that many survivors are taking health information they find on the Internet to discuss with their physicians. Although we found reports of physicians participating in online clinical discussion groups [52], we were unable to find studies of physician participation in online illness support groups with survivors and caregivers. Across all lists in our study, the median number of messages from identifiable health care providers was 1%. In five groups, no messages had senders identifiable as health care professionals. The prostate list had the largest proportion of messages sent by health care providers (4%). Further research is needed to identify barriers to participation by health care providers as well as ways to increase the motivation of health professionals to become active in online support groups. Potential barriers include health care providers' overloaded practice schedules, lack of incentives to seek information about online support resources, and potential liability issues associated with offering medical advice or opinions online.

External Validity

What does external validity mean in the case of naturally forming peer to-peer groups on the Internet, such as those in ACOR? Here we define it in terms of what new list members can realistically expect to find when they join a group. Early research suggests the pattern of predominance of offers of support over requests that we found among participants on ACOR lists is similar to those in other groups [21,50,51]. However, the pattern of messages containing much more informational support than emotional support probably reflects the highly technical nature of cancer treatments. Further research is needed to determine whether online groups for people living with chronic health and mental health conditions focus less on treatment technologies and more on emotional support for coping.

ACOR's organizational structure is another factor that could limit whether such health-related ecommunities can be replicated. ACOR's mailing list for listowners (Oncolist) appears to play a key role in developing and sustaining organizational capacity. Oncolist serves as a forum for listowners' discussions on the technical management of ACOR's Listserv software, virtual organizational development, and discussion facilitation as they relate to cancer-affected people. Listowners report that they value the information and support they receive from the list and read messages from the list regularly. Oncolist supports the long-term viability of all ACOR mailing lists by providing ongoing training for new listowners and encouragement for experienced ones. Through the list, listowners are also alerted when a listowner's deteriorating health or decision to step down precipitates a staffing problem on another listowner team. Some listowners report that they participate on multiple ACOR listowner teams, sharing their experience and reducing other listowners' workloads and the risk that a list will dissolve due to loss of leadership.

Other major Internet providers that host online communities have already recognized the need for such groups. Both Yahoo Groups and Google Groups host lists called "EL-M" (ElectronicMailingList-Manager) for listowners. However, these EL-M lists are open to an extremely heterogeneous collection of lists on widely divergent topics. Apart from our own preliminary unpublished studies, we could find no reports of research that tracked the organizational development of health eCommunities over time. In this context, it remains to be seen whether analogs to Oncolist will develop on Yahoo, Google, and elsewhere on the Internet to address the specialized needs of listowners who oversee ESGs for related health conditions.

Limitations

Email communication on mailing lists is fragmented. List members vary widely in how often they check their email and how often they post to lists. In lists with high volume, members can easily miss messages sent in reply to their posts. By systematically sampling the sequence of messages within list archives, we further fragmented the sequence and context of member correspondence. Although we expected that this sampling technique would, on balance, tap into correspondence of the most active members and identify the most common themes, we did not have the resources to confirm our interpretations with the full data set. Even though we sampled messages over a five-month period, our findings are summative. Discontinuities caused by our sampling techniques also prevented us from tracking within-group thematic changes over time. These situations are prime examples of the critical need for automated language analysis tools to examine individual sender behaviors within the rich context of online discussions as they occur over time.

Our analyses of participant characteristics revealed that caregivers were important participants in ACOR communities. It is beyond the scope of this article to present the perspectives of caregivers, as givers and recipients of support, but we will report on this topic in the future.

Finally, the mass of complex textual data produced by this study pushed the limits of conventional qualitative data analysis methods. To compare 10 lists in a single report, we necessarily omitted finer-grained descriptions of phenomena within individual lists. Our systematic sampling strategy constituted a trade-off. It supported this kind of macro analysis but made it impossible to ascertain whether messages that we used as the basis of identifying themes really represented dominant threads within the groups or were artifacts of the sampling process. The

scale of the study also limited the extent to which we could let list members "speak" for themselves through their email texts. Given the space allowed in most journal articles and the number of themes identified, we were limited to summarizing texts and offering exemplars. As with the demand for new analytic technologies, large-scale studies of Internet communities may also require new formats to document the discoveries we make about them.

This project used procedures to de-identify messages and perform systematic sampling to obtain data for each month in a timely manner. However, even using a sophisticated qualitative data analysis program like Atlas.ti to analyze messages, it took months to complete analyses because of the volume of messages and the number of coding categories we used. This slow process of data analysis contrasts sharply with the speed of the online communications that produced data. If research is going to accommodate the scale and scope of burgeoning eHealth support resources like mailing lists, finding ways to automate analytic processes will be essential. Information and social scientists, such as Pennebaker [62], Arguello [63], Kraut [personal communication, January 15, 2006], and Seale [64], have been working independently to develop efficient text data mining procedures that will work with semi-structured textual data, such as email messages [65,66]. As new technologies emerge, researchers will be able to study mailing list member interactions more comprehensively, over longer periods, and using a variety of methods. Thus, these innovative tools are critical to our capacity for rapidly gaining insights into the processes and discussion content of these groups. Ideally, in the future, researchers will use mixed methods—qualitative methods to explore the depth and richness of qualitative data and quantitative methods to achieve the potential precision and generalizability of quantitative data.

Conclusions

We examined the ways in which 10 cancer lists provide informational and emotional support. We found that, across all lists, cancer survivor participants were more likely to offer support than ask for it. Survivors communicated support primarily through offers of technical information and explicit advice about how to communicate with health care providers to get optimal care. Survivors offered support implicitly through accounts of their illness and treatment experiences. Explicit emotional support was less frequent than informational support and was often embedded in message texts. Emotional support often encouraged active coping, providing opportunities for cancer survivors to play empowered "helper" roles that reinforce self-esteem. Internet health support seekers report they are looking for both information and emotional support. This study offers strong evidence that they can and do find the support they need—and can, potentially, benefit from opportunities to play empowering roles in their online communities.

Acknowledgments

This research was funded by Robert Wood Johnson Foundation (RWJF) Health e-Technologies Initiative grant # 49152. This report does not represent the views of the Foundation. The authors would like to thank Kurt Ribsl, co-investigator on the UNC team, for his help in the conceptualization stage of this paper. We also thank Charles Rawls, Chief Technical Officer of ACOR, and the listowners of the participating ACOR lists. This research could not have been completed without their assistance. We also thank Lari Wentzel, PhD, University of California at Irvine, for her contributions to our understanding of stress and coping.

Authors' Contributions

ACOR's founder, Gilles Frydman, initiated the UNC-ACOR collaboration and was an essential contributor to project development and implementation. Frydman performed the programming required to extract and de-identify email messages from the ACOR archives for the qualitative study and provided extensive background information on the inner workings of ACOR and its constituent lists. In addition, he provided background data on member participation rates for the 10 lists over the data collection period.

Conflicts of Interest

None declared.

References

1. Fox S. Health information online: eight in ten Internet users have looked for health information online, with increased interest in diet, fitness, drugs, health insurance, experimental treatments, and particular doctors and hospitals. Washington, DC: Pew Internet & American Life Project; 2005. URL: http://www.pewinternet.org/pdfs/PIP_Healthtopics_May05.pdf [accessed 2005 Nov 15] [WebCite Cache ID 5I9svz1Vb]
2. Lenhart A, Horrigan J, Allen K, Boyce A, Madden M, O'Grady E. The ever-shifting Internet population: a new look at Internet access and the digital divide. Washington, DC: Pew Internet & American Life Project; 2003. URL: http://www.pewinternet.org/pdfs/PIP_Shifting_Net_Pop_Report.pdf [accessed 2003 Jul 12] [WebCite Cache ID 5I9tAaH1u]
3. Meric F, Bernstam EV, Mirza NQ, Hunt KK, Ames FC, Ross MI, et al. Breast cancer on the world wide web: cross sectional survey of quality of information and popularity of websites. BMJ 2002 Mar 9;324(7337):577-581 [FREE Full text] [Medline: 11884322] [doi: 10.1136/bmj.324.7337.577]

4. Dutta-Bergman MJ. Health attitudes, health cognitions, and health behaviors among Internet health information seekers: population-based survey. J Med Internet Res 2004 May 28;6(2):e15 [FREE Full text] [Medline: 15249264] [doi: 10.2196/jmir.6.2.e15]
5. Ybarra ML, Suman M. Help seeking behavior and the Internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [Medline: 16129659] [doi: 10.1016/j.ijmedinf.2005.07.029]
6. Rainie L. Who uses the Internet, what they do, and what it means. Washington, DC: Pew Internet & American Life Project; 2005. URL: http://www.pewinternet.org/ppt/Freedom%20to%20Connect%20Conference%20Speech.pdf [accessed 2005 May 30] [WebCite Cache ID 5I9tmkSJi]
7. Rheingold H. The Virtual Community. 1998. URL: http://www.rheingold.com/vc/book/ [accessed 2005 Jan 13] [WebCite Cache ID 5N4RkFMxU]
8. Levy JA, Strombeck R. Health benefits and risks of the Internet. J Med Syst 2002 Dec;26(6):495-510. [Medline: 12385532] [doi: 10.1023/A:1020288508362]
9. Horrigan J, Rainie L. Getting serious online. Washington, DC: Pew Internet & American Life Project; 2002. URL: http://www.pewinternet.org/pdfs/PIP_Getting_Serious_Online3ng.pdf [accessed 2006 Jan 11] [WebCite Cache ID 5I9uCU533]
10. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [Medline: 15142921] [doi: 10.1136/bmj.328.7449.1166]
11. Kummervold PE, Gammon D, Bergvik S, Johnsen JAK, Hasvold T, Rosenvinge JH. Social support in a wired world: use of online mental health forums in Norway. Nord J Psychiatry 2002;56(1):59-65. [Medline: 11869468] [doi: 10.1080/08039480252803945]
12. Lieberman MA, Russo S. Self help groups and the Internet: breast cancer newsgroups. Int J Self Help Self Care 2001;1(4):323-344. [doi: 10.2190/EGWJ-GBQD-KPDA-7UUD]
13. Steinberg DM. The Mutual Aid Approach to Working with Groups: Helping People Help One Another. 2nd ed. Binghamton, NY: Haworth Press; 2004.
14. Nonnecke B, Preece J. Lurker demographics: counting the silent. Proceedings of CHI 2000; 2000 April 1-6; Hague, The Netherlands. 2000. URL: http://www.ifsm.umbc.edu/~preece/paper/13%20Nonnecke&PreeceCHIpaper_ACMnotice.PDF [accessed 2006 Aug 15] [WebCite Cache ID 5I9ueRCJO]
15. Durning AB. Action at the grassroots: fighting poverty and environmental decline. Worldwatch Paper. 1989. p. 1-70 URL: http://www.uri.edu/mind/Reforming_%20Development_%20Assistance2.pdf [WebCite Cache ID 5OpoiZFHX]
16. Fisher R. Social action community organization: proliferation, persistence, roots, and prospects. In: Minkler M, editor. Community Organizing & Community Building for Health. New Brunswick, NJ: Rutgers University Press; 1997:53-65.
17. Klemm P, Bunnell D, Cullen M, Soneji R, Gibbons P, Holecek A. Online cancer support groups: a review of the research literature. Comput Inform Nurs 2003 May;21(3):136-142. [Medline: 12792194] [doi: 10.1097/00024665-200305000-00010]
18. Osheroff JA. Online, health-related discussion groups: what we should know and do. J Gen Intern Med 1997 Aug;12(8):511-512 [FREE Full text] [Medline: 9276659]
19. Glasgow RE, Boles SM, Mckay HG, Feil EG, Barrera M. The D-Net diabetes self-management program: long-term implementation, outcomes, and generalization results. Prev Med 2003 Apr;36(4):410-419. [Medline: 12649049] [doi: 10.1016/S0091-7435(02)00056-7]
20. White M, Dorman SM. Receiving social support online: implications for health education. Health Educ Res 2001 Dec;16(6):693-707 [FREE Full text] [Medline: 11780708] [doi: 10.1093/her/16.6.693]
21. Klaw E, Huebsch P, Humphreys K. Communication patterns in an on-line mutual help group for problem drinkers. J Community Psychol 2000;28(5):535-546. [doi: 10.1002/1520-6629(200009)28:5<535::AID-JCOP />3.0.CO;2-0]
22. Salem DA, Bogat GA, Reid C. Mutual help goes on-line. J Community Psychol 1997;25(2):189-207. [doi: 10.1002/(SICI)1520-6629(199703)25:2<189::AID-JCOP />3.0.CO;2-T]
23. 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 Apr;5(2):135-159. [Medline: 11010346]
24. 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]
25. Ferguson T, Frydman G. The first generation of e-patients. BMJ 2004 May 15;328(7449):1148-1149. [Medline: 15142894] [doi: 10.1136/bmj.328.7449.1148]
26. Dunham PJ, Hurshman A, Litwin E, Gusella J, Ellsworth C, Dodd PW. Computer-mediated social support: single young mothers as a model system. Am J Community Psychol 1998 Apr;26(2):281-306. [Medline: 9693693] [doi: 10.1023/A:1022132720104]
27. Houston TK, Cooper LA, Ford DE. Internet support groups for depression: a 1-year prospective cohort study. Am J Psychiatry 2002 Dec;159(12):2062-2068 [FREE Full text] [Medline: 12450957] [doi: 10.1176/appi.ajp.159.12.2062]
28. Bass DM, Mcclendon MJ, Brennan PF, Mccarthy C. The buffering effect of a computer support network on caregiver strain. J Aging Health 1998 Feb;10(1):20-43. [Medline: 10182416] [doi: 10.1177/089826439801000102]

29. Barrera M, Glasgow RE, Mckay HG, Boles SM, Feil EG. Do Internet-based support interventions change perceptions of social support?: An experimental trial of approaches for supporting diabetes self-management. Am J Community Psychol 2002 Oct;30(5):637-654. [Medline: 12188054] [doi: 10.1023/A:1016369114780]
30. Rimer BK, Lyons EJ, Ribisl KM, Bowling JM, Golin CE, Forlenza MJ, et al. How new subscribers use cancer-related online mailing lists. J Med Internet Res 2005 Jul 1;7(3):e32 [FREE Full text] [Medline: 15998623] [doi: 10.2196/jmir.7.3.e32]
31. Penedo FJ, Gonzalez JS, Davis C, Dahn J, Antoni MH, Ironson G, et al. Coping and psychological distress among symptomatic HIV+ men who have sex with men. Ann Behav Med 2003 Dec;25(3):203-213. [Medline: 12763715] [doi: 10.1207/S15324796ABM2503_06]
32. Wenzel L, Glanz K, Lerman C. Stress, coping and health behavior. In: Glanz K, Rimer BK, Lewis FM, editors. Health Behavior and Health Education: Theory, Research, and Practice. 3rd edition. San Francisco: Jossey-Bass; 2002:210-239.
33. Wenzel LB, Donnelly JP, Fowler JM, Habbal R, Taylor TH, Aziz N, et al. Resilience, reflection, and residual stress in ovarian cancer survivorship: a gynecologic oncology group study. Psychooncology 2002 Mar;11(2):142-153. [Medline: 11921330]
34. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York: Springer; 1994.
35. Brennan J. Adjustment to cancer - coping or personal transition? Psychooncology 2001 Jan;10(1):1-18. [Medline: 11180573]
36. Fleishman JA, Sherbourne CD, Crystal S, Collins RL, Marshall GN, Kelly M, et al. Coping, conflictual social interactions, social support, and mood among HIV-infected persons. HCSUS Consortium. Am J Community Psychol 2000 Aug;28(4):421-453. [Medline: 10965385] [doi: 10.1023/A:1005132430171]
37. Muresan S, Tzoukermann E, Klavans JL. Combining linguistic and machine learning techniques for email summarization. Proceedings of Conference on Computational Natural Language Learning (CoNLL-2001);. 2001 Jul 6. URL: http://www1.cs.columbia.edu/~smara/publications/muresan-conll2001.pdf [accessed 2006 Aug 15] [WebCite Cache ID 5I9wLLJbB]
38. Butler BS. Membership size, communication activity and sustainability: a resource-based model of online social structures. Information Systems Res 2001;12(4):346-362. [doi: 10.1287/isre.12.4.346.9703]
39. Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M. Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. Am Psychol 2004 Feb;59(2):105-117. [Medline: 14992637] [doi: 10.1037/0003-066X.59.2.105]
40. Atlas-TI [computer program]. Version 5.. Berlin, Germany; 2005.
41. Layder D. Sociological Practice: Linking Theory and Social Research. London, UK: Sage Publications; 1998.
42. Glaser B, Strauss A. The Discovery of Grounded Theory. Chicago, IL: Aldine Publishing Co; 1967.
43. Glaser B. Theoretical Sensitivity: Advances in the Methodology of Grounded Theory. Mill Valley, CA: Sociology Press; 1978.
44. Miles MB, Huberman AM. Qualitative Data Analysis: An Expanded Sourcebook. Thousand Oaks, CA: Sage; 1994.
45. ; National Cancer Institute. SEER Cancer Statistics Review, 1975-2002. URL: http://seer.cancer.gov/csr/1975_2002/ [accessed 2005 Nov 16] [WebCite Cache ID 5I9wk8Sqb]
46. Walther JB. Computer-mediated communication: Impersonal, interpersonal and hyper-personal interaction. Commun Res 1996;23(1):3-43. [doi: 10.1177/009365096023001001]
47. Cutrona CE, Suhr JA. Social support communication in the context of marriage: an analysis of couples' supportive interactions. In: Burleson BR, Albrecht TL, Sarason IG, editors. Communication of Social Support: Messages, Interactions, Relationships, and Community. Newbury, CA: Sage Publications; 1994:113-135.
48. Levenson RW, Ruef AM. Empathy: a physiological substrate. J Pers Soc Psychol 1992 Aug;63(2):234-246. [Medline: 1403614]
49. Preece J. Empathic communities: reaching out across the web. Interactions 1998;2(2):32-43. [doi: 10.1145/274430.274435]
50. Lasker JN, Sogolow ED, Sharim RR. The role of an online community for people with a rare disease: content analysis of messages posted on a primary biliary cirrhosis mailinglist. J Med Internet Res 2005 Mar 31;7(1):e10 [FREE Full text] [Medline: 15829472] [doi: 10.2196/jmir.7.1.e10]
51. Schoch N, White MD. A study of communication patterns of participants in consumer health electronic discussion groups. Proceedings of the 60th ASIS Annual Meeting 1997:November 1-6; Washington, DC.
52. Angelo SJ, Citkowitz E. An electronic survey of physicians using online clinical discussion groups: a brief report. Conn Med 2001 Mar;65(3):135-139. [Medline: 11291565]
53. Nonnecke B, Preece J. Why lurkers lurk. Proceedings of Americas Conference on Information Systems 2001;. 2001 Aug 3. URL: http://www.ifsm.umbc.edu/~preece/Papers/AMCISlurker.01.pdf [accessed 2006 Aug 15] [WebCite Cache ID 5I9xEG7Ip]
54. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the Internet affects patients' experience of cancer: a qualitative study. BMJ 2004 Mar 6;328(7439):564. [Medline: 15001506] [doi: 10.1136/bmj.328.7439.564]
55. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [Medline: 8555924]
56. Culver JD, Gerr F, Frumkin H. Medical information on the Internet: A study of an electronic bulletin board. J Gen Intern Med 1997 Aug;12(8):466-470. [Medline: 9276651] [doi: 10.1046/j.1525-1497.1997.00084.x]

57. Mursch K, Behnke-Mursch J. Internet-based interaction among brain tumour patients. Analysis of a medical mailing list. Zentralbl Neurochir 2003;64(2):71-75. [Medline: 12838475] [doi: 10.1055/s-2003-40375]
58. White MD. Questioning behavior on a consumer health electronic list. Library Q 2000;70(3):302-334.
59. Gerber BS, Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. J Med Internet Res 2001 Apr 9;3(2):E15 [FREE Full text] [Medline: 11720957] [doi: 10.2196/jmir.3.2.e15]
60. Eysenbach G, Till JE. Ethical issues in qualitative research on Internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [Medline: 11701577]
61. Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the Internet on health care and the physician-patient relationship: national U.S. survey among 1.050 U.S. physicians. J Med Internet Res 2003 Aug 29;5(3):e17 [FREE Full text] [Medline: 14517108] [doi: 10.2196/jmir.5.3.e17]
62. Pennebaker JW, Booth RJ, Francis ME. Linguistic Inquiry and Word Count (LIWC): A computerized text analysis program [computer program]. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.
63. Arguello J, Butler BS, Joyce E, Kraut R, Ling KS, Wang X. Talk to me. Foundations for successful individual-group interactions in online communities. Proceedings of CHI 2006, April 24-27. Quebec, Canada; 2006. URL: http://www.cs.cmu.edu/~kraut/RKraut.site.files/articles/arguello06-Foundations%20for%20Success%20_final.pdf [accessed 15 Aug 2006] [WebCite Cache ID 5I9y8UEd9]
64. Seale C, Ziebland S, Charteris-Black J. Gender, cancer experience and Internet use: a comparative keyword analysis of interviews and online cancer support groups. Soc Sci Med 2006 May;62(10):2577-2590. [Medline: 16361029] [doi: 10.1016/j.socscimed.2005.11.016]
65. Dorre J, Gerstl P, Seifert R. Text data mining: finding nuggets in mountains of textual data. In: Bramer MA, editor. Knowledge Discovery and Data Mining. London, UK: The Institution of Electrical Engineers; 1999:398-401.
66. Lagus K, Honkela T, Kaski S, Khonen T. WEBSOM for textual data mining. Artificial Intelligence Review 1999;13(5/6):345-364. [doi: 10.1023/A:1006586221250]

Abbreviations

ACOR: Association of Cancer Online Resources
CLL: chronic lymphocytic leukemia
ESG: electronic support group
UNC: University of North Carolina at Chapel Hill

submitted 01.09.06; peer-reviewed by H Winefield; comments to author 27.11.06; revised version received 20.02.07; accepted 20.02.07; published 14.05.07.

Please cite as:
Meier A, Lyons EJ, Frydman G, Forlenza M, Rimer BK
How Cancer Survivors Provide Support on Cancer-Related Internet Mailing Lists
J Med Internet Res 2007;9(2):e12
URL: http://www.jmir.org/2007/2/e12/
doi:10.2196/jmir.9.2.e12
PMID:17513283

Original Paper

User-Centered Research on Breast Cancer Patient Needs and Preferences of an Internet-Based Clinical Trial Matching System

Nancy L Atkinson[1], PhD; Holly A Massett[2], PhD; Christy Mylks[2], BA; Bethany Hanna[3], MS; Mary Jo Deering[4], PhD; Bradford W Hesse[5], PhD

[1]Department of Public and Community Health, University of Maryland, College Park, MD, USA

[2]Operations Research Office, National Cancer Institute, NIH, Rockville, MD, USA

[3]School Nutrition Association, Alexandria, VA, USA

[4]NCI Center for Bioinformatics, National Cancer Institute, NIH, Rockville, MD, USA

[5]Health Communications and Informatics Research Branch, National Cancer Institute, NIH, Rockville, MD, USA

Corresponding Author:
Nancy L Atkinson, PhD
University of Maryland
Department of Public and Community Health
Suite 2387 Valley Drive
College Park, MD 20742
USA
Phone: +1 301 405 2522
Fax: +1 301 314 9167
Email: atkinson@umd.edu

Abstract

Background: Internet-based clinical trial matching systems have the potential to streamline the search process for women with breast cancer seeking alternative treatments. A prototype system was developed to leverage the capabilities of a personal health record system for the purpose of identifying clinical trials.

Objective: This study examines how breast cancer patients perceive and interact with a preliminary version of an Internet-based clinical trial matching system, while taking into account the demands of diagnosis and treatment decision making.

Methods: Breast cancer patients participated in small group discussions and interacted with the prototype website in a two-phase qualitative research process. The first phase explored the experience of breast cancer patients (n = 8) with treatment decision making, initial responses to the idea of Internet-based clinical trial matching systems, and reactions to the prototype site. In the second phase, a different set of breast cancer patients (n = 7) reviewed revised website content and presentation and participated in a usability test in which they registered on the system and completed a personal health record to set up the matching process.

Results: Participants were initially skeptical of the prototype system because it emphasized registration, had a complicated registration process, and asked for complex medical information. Changing content and attending to usability guidelines improved the experience for women in the second phase of the research and enabled the identification of functionality and content issues, such as lack of clear information and directions on how to use the system.

Conclusions: This study showed that women felt favorably about the idea of using the Internet to search for clinical trials but that such a system needed to meet their expectations for credibility and privacy and be sensitive to their situation. Developers can meet these expectations by conforming to established usability guidelines and testing improvements with breast cancer patients. Future research is needed to verify these findings and to continue to improve systems of this nature.

(J Med Internet Res 2007;9(2):e13) doi:10.2196/jmir.9.2.e13

KEYWORDS

Software design; user-computer interface; breast neoplasms; personal health records; qualitative research; user-centered design; human-computer interaction

Introduction

Despite efforts to improve recruitment to cancer clinical trials, cancer patients continue to have difficulty understanding the research process and finding trials that they are interested in and qualify for [1]. Communications technology may be able to facilitate the clinical trial search process [2,3]. Already, the Internet provides an accepted tool for health information seeking [4] and clinical trial enrollment [3], and its use continues to spread widely throughout the United States [5]. Electronic personal health record (PHR) systems, another emerging technology, have the potential to further empower patients by enabling them to collect, store, and share their personal health information [6]. This paper explores the efficacy of enhancing an Internet-based prototype clinical trial matching system with the types of information collected in electronic PHRs to help women with breast cancer locate clinical trials most appropriate for their health condition.

Clinical trial enrollment is necessary to conduct the trials that advance medical research and practice [7]. Patients make the decision to participate in clinical trials for a variety of reasons, including the desire to help in the development of new and improved medical treatments [8]. Some participate out of a sense of altruism and wanting to help others who may be coping with a particular health problem [9,10]. Other patients seek trials that provide access to promising new treatments, which is particularly important if standard treatments have proven ineffective for them [10].

Nonparticipation in clinical trials is well documented. Many people are simply unaware of them or are uncomfortable with experimentation or participating in a trial with unknown outcomes; others fear losing health insurance or control of their treatment [9-12]. The clinical trial search process serves as another barrier [9-11]. In order to find trials, patients must be versed in the medical terminology related to cancer diagnosis and treatment and must understand the clinical trial research process in general and the demands of trials [11]. For example, patients who do not understand randomization or the different phases of clinical trial research will have trouble identifying a trial they might prefer or even asking questions about the trial [11]. Finally, the nature of the trial protocol is a barrier to enrollment [10,11]. Patients must meet certain inclusion and exclusion criteria in order to qualify for a trial; specifically, they must match the profile of a certain type of patient for whom the treatment is intended. As a result, patients have often been frustrated in their attempts to identify, qualify for, and join a trial [13].

National efforts are being made to empower patients through access and control of their medical records by maintaining a PHR [6]. The PHR is “an electronic application through which individuals can access, manage, and share their health information in a secure and confidential environment” [14]. PHRs have the potential to place the same information that clinicians have in the hands of patients so they can make informed decisions about their health care. A recent survey showed that the general public perceives that PHRs would improve their understanding of health conditions and their sense of control [15]. For the purpose of finding clinical trials, a detailed PHR-modeled system may also have the potential to facilitate a productive search by matching specific medical history information to the inclusion and exclusion criteria of relevant trials [3].

The purpose of the current study was to assess the usefulness, appeal, and functionality of an Internet-based clinical trial matching program from the perspective of the intended users: breast cancer patients. BreastCancerTrials.org (BCT) was the prototype system used in the current research study. This Internet-based system was initiated by two breast cancer survivors, working with the University of California, San Francisco (UCSF) Comprehensive Cancer Center and the Center of Excellence for Breast Cancer Care. BCT was subsequently developed as a working tool by the Center for Bioinformatics at the National Cancer Institute (NCI), in collaboration with UCSF and NCI’s Office of Communications, which manages NCI’s clinical trials database, PDQ. The site was designed to match patient medical histories to the eligibility criteria of participating clinical trials via a patient-created online medical record similar to a PHR that captures detailed information about current health, diagnosis, and treatment. Unlike most clinical trial search tools, BCT narrows the list of clinical trials based on the person’s actual medical record and seeks to limit frustration by improving the chance of a productive search.

Most services to search clinical trials provide listings of trials that may or may not allow users to filter results. More elaborate decision support tools such as BCT [16,17] have been developed for clinicians and investigators to aid in identifying potential research participants. A previous study on one Internet-based matching system for patients [3] showed that people were willing to use such a system and were able to enroll in trials. However, the findings were based on the people who actually found a match rather than all the users of the system [3], so the study did not reveal the reactions of users overall or any issues related to the experience of using the system.

Little other research has been conducted on Web-based search tools for cancer clinical trials. One study looked at cancer clinical trial websites that could be located from popular Internet search engines [18] and found that almost half of the 66 sites that were identified offered some kind of clinical trial search function. Researchers concluded that clinical trial websites had promise for increasing clinical trial recruitment, but barriers still existed that would hinder their use, such as large amounts of complex information and the lack of confidentiality. These findings were confirmed in a study [19] of the clinical trial search tools on the websites of NCI Comprehensive Cancer Centers. Most search mechanisms allowed the user to search by type of cancer but used medical terminology rather than lay language (eg, melanoma instead of skin cancer), and search results were found to be written at an eleventh grade level according to a readability analysis.

Because using BCT depends on having patients register and accurately complete the entire medical record for successful matching to clinical trials, research was necessary to assess how patients perceived, interacted with, and understood the site. This research also sought to establish how such a system fits into the

experience of breast cancer patients when making treatment decisions and to assess whether use of the system posed an additional barrier [20]. Since many breast cancer patients are likely to be older, and older adults' ability to use the Internet differs from other populations, the research also needed to examine interface design issues that may arise due to the demographics of the target population [21]. Three research questions guided the study:

1. How do breast cancer patients make treatment decisions in general and when presented with the prospect of an Internet-based clinical trial matching system?
2. What is the perceived usefulness of a prototype online system (ie, BCT) for matching breast cancer patients to clinical trials?
3. In what ways do the system's design and functionality help or hinder the clinical trial matching process?

Methods

Research Design

This qualitative study was conducted in March 2005 in two phases. In the first phase, researchers focused on general breast cancer diagnosis and treatment experiences, reactions to clinical trial participation, and reactions to the general idea of an Internet-based clinical trial matching system and to the prototype system. Based on the results of the first phase, the introductory site content and approach (home page, etc) were revised (see the Multimedia Appendix for before and after versions of the home page). The second phase of the study was a usability test of the site that focused on participant reactions to the revised introductory materials and to their experience of completing a PHR and finding clinical trials.

The use of two phases allowed thorough discussion and exploration of the issues and system while also not overburdening the respondents. This strategy also allowed the researchers to assess whether the revised content was more effective than the original version in orienting users to the prototype system, which is consistent with the usability guideline of evaluating websites before and after making changes [20].

A key component of the research design was the use of triads, in which three women participated in group discussion with a moderator. Like focus groups, triads provide an opportunity to explore issues in depth but allow each person to convey more information in a session [22]. The smaller group size, however, enabled the researchers to provide a more intimate environment for the participants who were coping with a sensitive health problem.

Sample

For the purposes of this research, NCI hired a professional recruitment firm to identify women eligible for the study from a marketing list they compiled on regular basis. The firm was provided a screener questionnaire to recruit a diverse mix of participants on the following criteria. All the participants in the study were female breast cancer patients age 18 or older living in the Washington, DC, metropolitan area. The study also sought to recruit patients who represented the full range of breast cancer diagnostic stages (stages 1 to 4) and patients with and without experience in clinical trials. Ethnic/racial diversity was requested but not required.

Women were excluded from participating in the study if they were employed by the US Department of Health and Human Services, the American Cancer Society, or other breast cancer foundations. Medical professionals, Web developers, information architects, and computer programmers were also excluded from participation because their technical knowledge and work experiences could influence their responses and perceptions. Women who did not use the Internet at least monthly were also excluded.

The first phase of the study included three triads of patients with stage 1 to 3 breast cancer who had not been in a clinical trial and one triad of women with stage 4 breast cancer who were currently in a clinical trial. The second phase consisted of one in-depth interview with a clinical trial patient with stage 4 breast cancer and two triads of patients with stage 1 to 3 breast cancer who had not been in a clinical trial. None of the participants from the first phase of the study participated in the second phase.

Instrumentation

Both study phases covered four topics: (1) personal experience with breast cancer, (2) perceptions of clinical trials, (3) the process of finding out about clinical trials, and (4) reaction to the prototype website. The questions in the first phase focused more heavily on the first three topics, and the second phase focused on topics 3 and 4 (see Table 1 for specific question areas). Phase 1 participants were asked more extensively about their experiences with breast cancer diagnosis, treatment, and information seeking and about their clinical trial perceptions, while phase 2 participants were asked to share primarily what they experienced at the time of diagnosis and any experience they had with clinical trials. Phase 2 participants spent more time reviewing the revisions that had been recommended in phase 1. Phase 1 participants did not examine the prototype in depth after completing the registration process, but phase 2 participants were asked to complete a medical record and try to find a trial.

Table 1. Discussion topics by research phase and category

Category	Phase 1 Topics	Phase 2 Topics
Personal Experience	Breast cancer diagnosis	Breast cancer diagnosis
	Finding out about treatment options	
	Ease or difficulty of finding information	
	Physician's role in treatment	
Perceptions of Clinical Trials	Knowledge of clinical trials	Personal experiences in clinical trials
	Positive and negative perceptions	
	Finding out about clinical trials	
	Perceptions of clinical trial participants	
	Personal experiences in clinical trials	
Exploring Clinical Trials	How they would learn about clinical trials	Personal experiences of learning about clinical trials, if any
	Knowledge and perceptions of online clinical trial tools	
Review of Prototype	Home page general impressions and acceptability	Review of revised home page text, graphics, and layout
	Ease of navigation	Observation of initial expectations and reactions to site
	Review of background material provided on the site	Ease of log-in process and perceived acceptance
	Review of privacy policy text	Ease of the PHR process and perceived acceptance
	General impressions and acceptability of information requested when completing the PHR/profile	Home page general impressions and acceptability
		Ease of navigation
		Reactions to logging in to the site
		Review of background material provided on the site
		Review of privacy policy and informed consent text
		General impressions and acceptability of information requested when completing the PHR/profile

Data Collection

The study took place in the usability testing facility of the User-Centered Informatics Research Laboratory in NCI's Operations Research Office. In both phases of the study, a skilled moderator not involved with website development conducted the interviews with the assistance of a usability specialist. Participants each sat at a computer with access to the BCT site so as to provide a real-time assessment of ease of use, desirability, and thought process. The iterative process first involved a group discussion of the first three topic areas, followed by a series of one-on-one computer interaction sessions with different parts of the website (eg, homepage, "About Clinical Trials" page), and then followed by group discussions reacting to each website part. Participants also were asked to provide written feedback on printed copies of the Web pages, making notes and highlighting sections they liked and disliked. These sessions were audiotaped and videotaped. Also, through a two-way mirror, research assistants observed the interviews, the women's use of the prototype system, and their nonverbal reactions to working with it.

As described above, each phase had different goals, which influenced the data collection process. The first phase assessed overall clinical trial attitudes and experiences as well as reactions to the existing site and introductory materials on the home page and subpages. After phase 1 was completed, a website designer revised the BCT site pages based on the feedback obtained and mocked up paper drafts of each revised page. During phase 2, participants took part in a usability test of the existing website in which they were asked to attempt to accomplish three tasks: register, fill in the medical record forms, and review matches. For each task, they were first observed using the site and were then given the opportunity to discuss their experiences as a group. Following an initial review of the site, participants reviewed the revised version and were asked to react as a group to the changes.

This study was determined to be exempted research because it was designed to evaluate consumer satisfaction for quality improvement rather than individual behavior. Nevertheless, only participants who agreed to participate and signed informed consent forms were included in the study. The data were anonymous and reported in the aggregate with no identifiers.

Data Analysis

The data were analyzed using content analysis, a qualitative method, via code mapping of transcripts [23]. Certain words, phrases, and quotes were grouped together based on similarity of themes regarding decision making, acceptability to the prototype website, and reactions to site design and functionality. Additional codes were created for topics that emerged beyond the initial codes. Similar concepts were summarized and formed into categories, which were evaluated for similarities and differences within and across groups. A grid was constructed to provide an overview summarizing the content of the discussions. Two primary reviewers were utilized to identify these themes, and two secondary reviewers examined the findings to ensure reliability and consensus across reviewers.

Results

A total of 15 women participated in the study, of which 8 participated in phase 1 and 7 in phase 2. Age of participants ranged from 36 to 78 years, and the average age was 54.8 (SD = 10.93). One third (n = 5) of the participants were African American; the rest were Caucasian. The findings below are separated into three sections highlighting key issues related to the research questions: (1) treatment decision making and finding clinical trials, (2) response to the prototype system for finding breast cancer clinical trials, and (3) site design and functionality.

Treatment Decision Making and Finding Clinical Trials

The results largely mirrored the findings of previous research that lack of awareness and concerns about experimentation prevented people from considering clinical trial enrollment. Women who had experience with clinical trials, however, were less skeptical about their usefulness and were more open to considering participating in the future. The responses also showed that clinicians could both facilitate and obstruct participation in clinical trials. Clinicians were the main source of information about treatment options, including clinical trials. All the respondents described having very little time between diagnosis and starting treatment to understand their disease or participate in their treatment decisions. Often their clinicians either recommended a course of action or made decisions as the women were under anesthesia during diagnostic procedures.

Participants' attitudes toward using a clinical trial matching tool were also influenced by knowledge and clinician support. Although most said they would use the Internet to seek information on clinical trials, most were unfamiliar with clinical trial matching tools. They would be more trusting of an Internet-based matching tool if a clinician recommended it to them, if it was made available in a clinician's office, or if it was provided by a trusted organization. Women who knew more about clinical trials were less concerned about using an Internet-based clinical trial matching tool than those with less knowledge.

Perceptions of Prototype Matching System

Phase 1 explored the initial reactions to the system and how it was portrayed to potential users on the home page and in the supporting text for the "About BCT" and the registration process. Table 2 presents the positive and negative reactions to the prototype system among the first phase respondents.

Table 2. Positive reactions and concerns to BCT prototype matching system among phase 1 participants

	Positive Reactions	Concerns
General Reactions	Participants had a favorable reaction to the overall concept and purpose of the site.	Users were confused by the acronyms (BCT, PHR, etc).
	They considered the idea to have much promise for women interested in finding a trial.	Site appeared more focused on the needs of researchers than of breast cancer patients.
		Many did not appreciate or respond to motivational quotes and art-work unrelated to the site's purpose, which seemed to add visual clutter but no substance.
Site Entry	Participants were pleased with use of the NCI logo, which improved the site's legitimacy.	Women wanted more visual cues that the site was focused on breast cancer.
		They wanted more information about the system and its benefits up front, before having to register.
		They wanted to see the types of trials available before registering.
		Participants wanted more information about the demands of trials before registering.
Learning About Clinical Trials	A link to "About Clinical Trials" was used often, especially by women who had little knowledge of clinical trials.	Information provided was for all cancer clinical trials, and women wanted targeted information for breast cancer trials.
Learning About the Site	A link to information about BCT was sought out first by women who had an understanding of clinical trials.	The site lacked clear information about the site's mission and purpose and sponsorship on the home page.
	Information on the secondary page was considered credible and informative.	Information on the secondary page seemed too long and did not sufficiently address benefits to patients.
Privacy Policy	Participants liked that a privacy policy was available.	Privacy policy seemed long and complex in its wording.
Registration		The emphasis on registration on the home page was off-putting.
		Many commented on the burden of constantly filling in medical forms.
		Participants preferred giving personal contact information only after getting matches.
Consent		Consent procedure was seen as long and cumbersome.
		Consent form seemed to conflict with privacy policy, saying that they risked having information accessed by a nonauthorized source.

Overall, the prototype matching system was seen favorably and as a meaningful innovation for breast cancer patients seeking treatment options. Women noticed information that increased their perceptions of the site's credibility and read about the sponsors' concerns about women with breast cancer. However, some women questioned the site's legitimacy; although not specifically asked to locate information about the site's sponsorship and mission, the women stated that they wanted to know about sponsorship and were concerned when it was not readily apparent.

When women were confused or felt pushed to register, their reactions became more negative, and they stated more reluctance to using the site. For example, participants had an immediate negative reaction to the prominence of the log-in area on the home page. They stated that they felt uncomfortable with having to register before they knew more about what the system could do for them and the benefits of registration. Feeling driven to register, women stated that the system was developed more for researchers than for breast cancer patients.

The registration process was viewed as cumbersome and reminded the women of filling out extensive medical forms related to their condition. In addition, registration included the need to review and agree to a lengthy consent form, which lengthened the process even more.

As a result of the first phase of triad interviews, several pages of the site were revised. The home page was revised with added information about the benefits of clinical trials, the benefits to using the prototype system, and information about privacy and confidentiality (see Multimedia Appendix). It also used a simple medical image rather than a larger decorative background image. A new page was created to give users who sought information about clinical trials a brief targeted overview of breast cancer clinical trials rather than linking directly to NCI's website for all clinical trials (ie, not just breast cancer trials), as was done in the original prototype. The page describing the BCT project was also revised to focus on patient benefits.

Participants in the second phase reviewed the revised pages and found them to be clearer and more assuring. Specifically, the users were less skeptical and more open to exploring the system. They seemed more reassured about the system and voiced no reluctance to completing the registration process, unlike the women reacting to the site in phase one.

In the second phase, participants were also asked to review the matches that they could receive after entering all of their information into the system. Women were provided with trial information from NCI's PDQ Cancer Clinical Trials Registry. The clinical trial records were very confusing and overwhelming to the participants because of the technical medical language. Women were particularly confused about the difference between stage of cancer and phase of clinical trial. Despite this confusion, seeing what trial information looked like helped the participants understand the system much better. Several women wanted to be able to see this type of information earlier in their interaction with the system. Although they liked that the system could tailor the types of trials to their profiles, they wanted to see all possible matches first and then have the ability to drill down themselves.

Effectiveness of Site Design and Functionality

The third research question of the study was about how the site's design and functionality affected the match process. These issues were touched on in the first phase of the study and were examined more thoroughly in the second phase as the women had more direct interaction with the site. Table 3 provides a summary of the key design and functionality concerns raised by women in phase 2.

Results demonstrated that users had difficulty using the system on entry, at registration, and when using and completing the PHR. On entry, users were confused about the relevance of the system to their needs. They had trouble locating information that conveyed the site's credibility, and, when they looked for more information, some of the links sent them to external Web pages without warning. Mostly, the emphasis on registration on the home page made them think they would be forced into joining a fee-based service before they knew what to expect. They were offended by having to provide personal contact information before they received information from the system. As participants proceeded to register on the system, the lack of clear and concise information in the privacy policy and consent form, the lack of clear directions, and poor error handling of the system hampered their progress further.

Once participants entered the personal home page and the PHR-like medical record, they had to orient themselves to new sections with different formats and functionality, and they often had difficulty understanding or using features. The section for describing the various treatments they had received required use of an add/edit/delete link that was not intuitive. Women found that they had trouble moving forward in other sections because of a lack of cues on how to successfully complete the section or what to do next.

Help text was useful but not always available, which was disconcerting for a few questions for which the implications of their answers were unclear. For example, a question asked about one's willingness to stop current treatment, and women were worried how their answer would affect their relationships with their clinicians or with researchers. Several women stated that they would want to talk to their clinicians before answering certain questions. Others were reluctant to answer questions, especially those asking for personal identifying information, because they wanted to know more about how the information would be used and what to expect in general.

Table 3. Design and functionality issues of prototype system for phase 2 participants

Section	Design and Functionality Issues	Illustrative Quotes
Home Page	Emphasis on registration on the home page was off-putting.	"If you have to register even to get past the home page...I'm not that committed to the site yet."
		"I think of when I have to [register] as places when I'm buying things."
	Women wanted design elements that demonstrated credibility and relevance to their needs.	"I'd go for the pink ribbon, because that lets you know it's a breast cancer site.... You know you're in the right place."
	External links to NCI's clinical trial information page caused users to get lost and even forget about returning to BCT.	"You kind of get over there [on the outside site] and have to think back to 'What was I originally doing?'.... I totally forgot I was supposed to be looking for clinical trials on this site [bct.org]. I started going deeper and deeper into the NCI site."
Registration	Users had trouble creating and remembering a username and password due to the confusing process.	"My initial reaction is, why do I have to register, and why do I have to remember another password?"
	No message or cue signaled when registration was successful.	"Is that it? Am I done?"
		"Where is the part that says I'm successful?"
Personal Home Page	Participants were confused about the purpose of this section compared to the overall home page and the medical record.	"I thought I was going back to the beginning [home]."
	The functionality of the menu items in this section was not clear, so users didn't know how or why they would use these materials.	"Why do I have an account [My Account]? Is that my personal health history, or my personal home page, or...?"
	The link to start the medical record was difficult to see unless they scrolled down the page—users did not know where to click.	
Uncomfortable Elements	Women were often upset when they were required to give personal information when they wanted to see what matching results would look like.	"If I just want to know where some trials are, why do they have to know really anything about me?"
	"My Cancer" was perceived as asking women to own their disease when they were in a battle to eliminate cancer from their bodies.	"If I leave nothing else with you, [please don't] call it 'My Cancer'.... That's a painful one... to see and have to answer questions on it...just brings you into a reality where you don't want to be when you're just trying to get some information.... We try not to take ownership...."
	Users wanted help from a clinician to know how to fill in technical information on diagnosis and treatment, especially when the implications of their responses were unclear.	
Confusing Elements	Users did not recognize the difference between the "Save and Remain on This Page" and the "Save and Go Forward" buttons.	"I'm looking for a way to go back.... It says do not use your browser's arrow buttons...[but the only available button is save and continue]."
	There was a lack of prompts on how to move forward or retrieve clinical trial matches after filling out the medical record.	"I'd look for something where it says that I've entered this information...."
		"Where does it tell me to save it? Or has it already saved it?"
Help Mechanisms	"Learn more" buttons were helpful when available. When context-specific help was not available, women struggled to respond.	
	Users had difficulty responding to error messages because of not being able to find the source of the problem or not knowing how to fix it.	"No, [I don't know why it's telling me the username is invalid]. Unless my name is just too obvious a name?"
	Too few instructions were available when completing the section about previous cancer treatments. They had difficulty with the add/edit/delete functions.	"It wasn't clear to me at first.... I'm thinking 'Surgery, what do they want? Yes, I've had it,' but when I saw Add, Delete, or Edit, how can I edit something that isn't there?"

Discussion

This study confirmed many of the findings concerning the design of usable websites and the barriers and motivators to clinical trial participation. It also added knowledge about the clinical trial search process in the context of an Internet-based tool. These findings will be useful to clinical trial researchers and Web professionals designing websites for adults coping with a chronic disease such as cancer.

Treatment and Clinical Trial Decision Making

In this study, we first examined the treatment decision-making process of individuals diagnosed with breast cancer. The findings confirmed that physicians, especially oncologists, were primary gatekeepers to information on which patients base their treatment plans [9]. Unlike other studies that have looked at specific barriers for cancer patients and breast cancer patients, this study also revealed that time constraints were another significant barrier to participating in clinical trials. Specifically, women newly diagnosed with breast cancer often felt caught in a whirlwind of daunting information with no time between diagnosis and treatment to better understand the disease, much less make treatment decisions. The participants described the breast cancer diagnosis and treatment process as time constrained, emotional, and as focused on following their medical providers' advice and treatment plans.

The findings also confirmed that women were universally in favor of clinical trials in concept [8,9], but most did not understand the actual process and the demands of participation [11]. The trial description and eligibility criteria were laden with medical terminology, making it difficult for most participants to understand and use the information in order to determine their options and preferences for participation [10]. This study found that time constraints limited the opportunity for patients to become familiar with clinical trials or consider them, as their health care providers were more focused on controlling and eradicating the cancer. These findings suggest that, in order to provide clinical trials as an alternative, health care providers and intermediaries need to be more involved in increasing awareness and knowledge about trials, specifically regarding their role in developing new treatments and the process, benefits, and risks of participation [24].

The study also sought to determine how the idea of a clinical trial search tool fit the needs and expectations of breast cancer patients. Most were unfamiliar with such tools, though a few vaguely remembered seeing online clinical trial search tools on the websites of NCI and the American Cancer Society. Most participants were interested in the concept of a PHR-type clinical trial matching tool. Women who had previous clinical trial experience had fewer concerns than those who had little knowledge of trials. Women said they would be more likely to try a matching service if clinicians recommended it or if the service was made available at the clinician's office where they could get help using it.

These results confirm previous research showing that cancer patients were willing to use a clinical trial matching tool, particularly if encouraged by health care providers, and that many relied on surrogates to help them [3]. The research suggests that intermediaries—family, clinicians, and clinical staff members—can help women access information on trials by introducing tools like the prototype system that enable them to find trials from which they may benefit.

Perceptions of Prototype

Our second research question addressed participant reactions to the prototype clinical trial matching system. Previous research found that referrals to clinical trials can be facilitated with technology for the clinician [17] and for patients [3]. Participants in the current study were also favorable to the concept of the prototype and thought it would be helpful to women looking for a breast cancer clinical trial.

This study also provided a unique view of breast cancer patients' experiences and reactions to the clinical trial search tool that will be valuable to product developers and clinical trial investigators. Looking at the website design recommendations developed by the US Department of Health and Human Services [20] and by the American Association of Retired Persons (AARP) [21], this study confirmed much of the evidence. In general, cues that demonstrated the credibility and legitimacy of the site, like the NCI logo, were well received, and the women searched for such information without being prompted. The women also wanted artwork and graphics that resonated for them—such as a pink ribbon graphic to designate it as a breast cancer site—or images that supported the messages of the site. They did not appreciate graphical elements that seemed unrelated to clinical trials, such as the presentation of artwork done by breast cancer survivors that appeared in the prototype system. These findings confirmed the evidence that relevant graphics were necessary to assure users that a site was credible [20]. Developers need to use pertinent graphical elements and feature the logos of sponsoring organizations with positive standing.

Participants needed more information about how the system could help them before they could develop an interest in it. This finding suggests that developers should provide more explanation about how such an Internet-based tool improves the typical way that women find clinical trials. One way to convey this information would be to make the purpose and benefits of the site easily understandable on the home page. For example, the site could allow women to browse the trials database or could provide a tour of the service. Another way to provide information would be to use second-level Web pages to convey the purpose and functionality of the site, present the benefits of using the site, and make frequently asked questions available. Both of these strategies were put into place for the second phase of the research, and the results received very positive reviews by the participants.

Communicating a website's purpose clearly on the home page is also a recommendation in the usability evidence base [20]. Although the usability evidence also recommends limiting the prose and length of the home page [20], the participants in this study appeared to favor more information rather than less on the home page as well as the secondary pages.

The findings also revealed that emphasizing and requiring registration on a clinical trial Internet site was a sensitive issue. Prior to being required to register, many participants wanted to understand what they would be asked to do (in terms of information disclosure), what they would get in return via the match process, and what the benefit to them would be. Many women reported being tired of the assumption that they would readily release their personal information to researchers in order to register. While the strategies mentioned above could go a long way toward orienting users, a tutorial or tour may further help potential users to fully understand how the site works, the

nature of their involvement, and what the results could be. Allowing users to browse trials before registering would also enable them to learn about the types of trials available and what they could expect to get out of engaging in the process.

Design and Functionality Issues

The third research question examined how the website's design and functionality affected the clinical trial match process. The current study confirmed findings in the usability literature about specific design features and functionality.

Usability research recommends that critical information be placed high in the hierarchy of the website and that information be organized clearly [20]. In the current study, the lack of knowledge about clinical trials was very common, and women sought information about clinical trails first. These findings support providing a link to site-specific clinical trial information ("About Clinical Trials") first among the options in the menu structure to help explain clinical trials upfront.

The study confirmed the usability recommendation to liberally use descriptive headings to help users conceptually relate to the content [20]. Participants, for example, had strong negative reactions to the label "My Cancer," believing the site asked them to "own" their cancer when they were trying to rid themselves of it. Therefore, these results also indicate that developers should pay careful attention to creating headings that are acceptable and sensitive to the target audience's situation.

The study showed that the opportunities for confusing potential users of this system proved to be great. The topic of clinical trials was difficult to present, in general, which is consistent with previous research [11]. Here, the women also found that the terminology related to breast cancer diagnosis and treatment and the privacy policy added to the overall complexity. The use of the Internet and informatics terms and acronyms like PHR further confused and overwhelmed the users trying to navigate the site. Limiting the use of acronyms and medical jargon and using terms that are more recognizable to the lay public are important in the content development of usable websites [20]. These findings support the placement of context-specific help text to address user questions at the point at which they are having trouble and to ensure ease of navigation. Users can also be prompted and guided on how to get assistance from their medical providers so they get the most out of the system.

The study also confirmed the recommendation to provide assistance to users, especially novices and first-time users [20]. Many breast cancer patients are likely to be older women, and older adults tend to have less experience and expertise with computers and the Internet [21]. As a result, participants often required additional help with registration tasks such as selecting a username and password. They were also more likely to require assistance when the system's navigational cues and input strategies varied because they did not know what they needed to do to enter information or move on in the program [21]. Therefore, the developers of such sites should strive to give clear directions and specific examples to facilitate use among the older adult target audience.

Previous usability studies have indicated that developers should indicate when a link will move users to a different location on the same page, a new page, or a different website [20]. The current study confirmed that external links were problematic. Specifically, users who clicked on the "About Clinical Trials" were confused when they went to the NCI website and had trouble getting back to the prototype site. Some even forgot about the prototype website and became engrossed in the information about clinical trials on the NCI site.

The evidence from usability research has indicated that action sequences should be developed so they are easily understood [20]. The tasks involved in the prototype system were complex in nature. This research suggests that users would benefit from revised instructions and additional prompts, especially when completing their PHR. For example, the section where users recorded medical treatments ("My Treatment") could be enhanced with better design or directions on how to add multiple entries (eg, if they had two lumpectomies in the same breast but at different times) or to delete entries. Once a task was completed, the system could also send a message to let participants know whether or not they successfully completed the task, such as registering on the site or completing the health record. If they did not successfully complete the task, the system should provide instructions of what they still need to do in order to be matched to clinical trials.

Conclusions and Recommendations

This study built upon previous research on clinical trial recruitment and usability, finding that even novice and older adult users were open to an Internet-based clinical trial matching system. In terms of breast cancer diagnosis and treatment, the system must also fit into the context of an emotionally charged, time-limited situation. Such a system must also fulfill expectations for credibility and benefit as well as privacy so that women feel confident and assured in their participation.

One limitation of the study was the use of a qualitative research design, which means that we cannot generalize the results to other breast cancer patients. Future research is needed that provides quantitative data and uses randomization. Another limitation was that the participants were not actually in the process of searching for a cancer clinical trial—they were only responding to a hypothetical situation. Therefore, their decision-making processes might have been different if they were truly looking for a clinical trial. Recruiting breast cancer patients for this study was challenging in general, but future research will be needed with women who are in the process of considering clinical trials. Lastly, the number of participants might be considered small by some researchers. Although the study may have benefited from increasing the sample size, user-centered research shows that as few as six people is sufficient.

The research confirmed one of the core tenets of usability engineering, that one must use an iterative design approach [20]. Checking with users and observing them with the site ensures relevance and usefulness of the resulting product. Although user-centered research takes additional time and effort, it saves time and money in the long run by identifying problems early.

Another primary approach in usability engineering is evaluating websites both before and after making changes. This was a first step in our research process. Future research is needed to confirm that changes made to the prototype system based on the user-centered research improve its usefulness and increase satisfaction among the target audience [20]. Qualitative research can be used to further examine user reactions. However, research could also be done to compare how easily and effectively users can navigate and use the prototype and modified websites. Most importantly, outcome research is needed to determine whether such tools actually result in more and better matches. Translational research would also enable exploration of how clinical trial matching systems could best be integrated into different settings, especially medical facilities and the home.

Acknowledgments

The National Cancer Institute's Operations Research Office funded study design; collection, management, analysis, and interpretation of the data; and preparation, review, and approval of the manuscript. The system was developed collaboratively by NCI's Center for Bioinformatics, which manages the cancer Biomedical Informatics Grid (caBIG) initiative; NCI's Office of Communications, which manages NCI's clinical trials database PDQ; and the University of California, San Francisco, Comprehensive Cancer Center and the UCSF Center of Excellence for Breast Cancer Care.

Conflicts of Interest

None declared.

Multimedia Appendix

Screenshots of the Web site (ppt) [PPT file (MS Powerpoint), 327 KB - jmir_v9i2e13_app1.ppt]

References

1. Sung NS, Crowley WF, Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. JAMA 2003 Mar 12;289(10):1278-1287. [Medline: 12633190] [doi: 10.1001/jama.289.10.1278]
2. Marks L, Power E. Using technology to address recruitment issues in the clinical trial process. Trends Biotechnol 2002;20:105-109. [Medline: 11841861] [doi: 10.1016/S0167-7799(02)01881-4]
3. Metz JM, Coyle C, Hudson C, Hampshire M. An Internet-based cancer clinical trials matching resource. J Med Internet Res 2005;7(3):e24 [FREE Full text] [Medline: 15998615] [doi: 10.2196/jmir.7.3.e24]
4. Fox S, Rainie L. Vital Decisions: How Internet Users Decide What Information to Trust When They or Their Loved Ones Are Sick. Washington, DC: Pew Internet & American Life Project; 2002. URL: http://www.pewinternet.org/pdfs/PIP_Vital_Decisions_May2002.pdf [WebCite Cache ID 5OpjXORPq]
5. ; National Telecommunications and Information Administration. A Nation Online: Entering the Broadband Age. Washington, DC: US Department of Commerce; 2004. URL: http://www.ntia.doc.gov/reports/anol/NationOnlineBroadband04.pdf [WebCite Cache ID 5OpjYfAhG]
6. ; Markle Foundation. Achieving Electronic Connectivity in Healthcare: A Preliminary Roadmap from the Nation's Public and Private-Sector Healthcare Leaders. New York: Markle Foundation; 2004. URL: http://www.connectingforhealth.org/resources/cfh_aech_roadmap_072004.pdf [WebCite Cache ID 5OpjZx8bL]
7. Nathan DG, Wilson JD. Clinical research and the NIH--a report card. N Engl J Med 2003 Nov 6;349(19):1860-1865. [Medline: 14602885] [doi: 10.1056/NEJMsb035066]
8. Lovato LC, Hill K, Hertert S, Hunninghake DB, Probstfield JL. Recruitment for controlled clinical trials: literature summary and annotated bibliography. Control Clin Trials 1997;18:328-352. [Medline: 9257072] [doi: 10.1016/S0197-2456(96)00236-X]
9. Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review. J Clin Epidemiol 1999 Dec;52(12):1143-1156. [Medline: 10580777] [doi: 10.1016/S0895-4356(99)00141-9]
10. Grunfeld E, Zitzelsberger L, Coristine M, Aspelund F. Barriers and facilitators to enrollment in cancer clinical trials: Qualitative study of the perspectives of clinical research associates. Cancer 2002;95(7):1577-1583. [Medline: 12237928] [doi: 10.1002/cncr.10862]
11. Mills EJ, Seely D, Rachlis B, Griffith L, Wu P, Wilson K, et al. Barriers to participation in clinical trials of cancer: a meta-analysis and systematic review of patient-reported factors. Lancet Oncol 2006 Feb;7(2):141-148. [Medline: 16455478] [doi: 10.1016/S1470-2045(06)70576-9]
12. Lara PN, Higdon R, Lim N, Kwan K, Tanaka M, Lau DH, et al. Prospective evaluation of cancer clinical trial accrual patterns: identifying potential barriers to enrollment. J Clin Oncol 2001 Mar 15;19(6):1728-1733 [FREE Full text] [Medline: 11251003]

13. Adams-Campbell LL, Ahaghotu C, Gaskins M, Dawkins FW, Smoot D, Polk OD, et al. Enrollment of African Americans into clinical treatment trials: Study design barriers. J Clin Oncol 2004;22:730-734 [FREE Full text] [Medline: 14966098] [doi: 10.1200/JCO.2004.03.160]
14. ; Markle Foundation. Connecting Americans to their Healthcare. Final Report: Working Group on Policies for Electronic Information Sharing between Doctors and Patients. New York: Markle Foundation; 2004. URL: http://www.connectingforhealth.org/resources/wg_eis_final_report_0704.pdf [WebCite Cache ID 5OpjigDv9]
15. Ross SE, Todd J, Moore LA, Beaty BL, Wittevrongel L, Lin C. Expectations of patients and physicians regarding patient-accessible medical records. J Med Internet Res 2005;7(2):e13 [FREE Full text] [Medline: 15914460] [doi: 10.2196/jmir.7.2.e13]
16. Breitfeld PP, Ullrich F, Anderson J, Crist WM. Web-based decision support for clinical trial eligibility determination in an international clinical trials network. Control Clin Trials 2003 Dec;24(6):702-710. [Medline: 14662275] [doi: 10.1016/S0197-2456(03)00069-2]
17. Embi PJ, Jain A, Clark J, Bizjack S, Hornung R, Harris CM. Effect of a clinical trial alert system on physician participation in trial recruitment. Arch Intern Med 2005 Oct 24;165(19):2272-2277 [FREE Full text] [Medline: 16246994] [doi: 10.1001/archinte.165.19.2272]
18. Simon C, Hegedus S. Exploring websites on cancer clinical trials: an empirical review. Contemp Clin Trials 2005;26(5):530-533. [Medline: 16122989] [doi: 10.1016/j.cct.2005.07.004]
19. Monaco V, Krills SK. On-line information about cancer clinical trials: evaluating the Web sites of comprehensive cancer centers. AMIA Annu Symp Proc 2003:470-474. [Medline: 14728217]
20. ; US Department of Health and Human Services. Research-Based Web Design & Usability Guidelines. Washington, DC: US Department of Health and Human Services; 2006. URL: http://usability.gov/pdfs/guidelines_book.pdf [WebCite Cache ID Oh9]
21. Redish J, Chisnell D. Designing Websites for Older Adults: A Recent Review of the Literature. Washington, DC: American Association of Retired Persons; 2004. URL: http://assets.aarp.org/www.aarp.org_/articles/research/oww/AARP-LitReview2004.pdf [WebCite Cache ID 5Opjp9hki]
22. Greenbaum TL. The Handbook for Focus Group Research. 2nd edition. Thousand Oaks, CA: Sage; 1998. URL: http://www.sagepub.co.uk/refbooksProdDesc.nav?prodId=Book6819
23. Morgan DL. Successful Focus Groups: Advancing the State of the Art. Newbury Park, CA: Sage; 1993. URL: http://www.sagepub.co.uk/booksProdDesc.nav?prodId=Book3927
24. Brawley O. The study of accrual to clinical trials: Can we learn from studying who enters our studies? J Clin Oncol 2004;22(11):2039-2040 [FREE Full text] [Medline: 15082727]

Abbreviations

BCT: BreastCancerTrials.org
NCI: National Cancer Institute
PHR: personal health record

submitted 27.11.06; peer-reviewed by P Embi, V Monaco; comments to author 21.12.06; revised version received 08.04.07; accepted 10.04.07; published 15.05.07.

Please cite as:
Atkinson NL, Massett HA, Mylks C, Hanna B, Deering MJ, Hesse BW
User-Centered Research on Breast Cancer Patient Needs and Preferences of an Internet-Based Clinical Trial Matching System
J Med Internet Res 2007;9(2):e13
URL: http://www.jmir.org/2007/2/e13/
doi:10.2196/jmir.9.2.e13
PMID:17513284

Original Paper

Pilot Randomized Trial of the Effect of Wireless Telemonitoring on Compliance and Treatment Efficacy in Obstructive Sleep Apnea

Carl J Stepnowsky Jr.[1,2], PhD; Joe J Palau[1], BA; Matthew R Marler[3], PhD; Allen L Gifford[4,5], MD

[1]Health Services Research & Development Service, Veterans Affairs San Diego Healthcare System (VASDHS), San Diego, CA, USA

[2]Department of Psychiatry, University of California, San Diego, CA, USA

[3]California Institute for Telecommunications and Information Technology (Calit2), University of California, Irvine, CA, USA

[4]VA Bedford Center for Health Quality, Outcomes & Economic Research, VA New England Healthcare, Bedford, MA, USA

[5]Department of Public Health and Medicine, Boston University, Boston, MA, USA

Corresponding Author:
Carl J Stepnowsky, PhD
VASDHS, Health Services Research & Development Service (111N-1)
3350 La Jolla Village Drive
San Diego, CA 92161
USA
Phone: +1 858 552 8585 ext 5948
Fax: +1 858 552 4321
Email: cstepnowsky@ucsd.edu

Abstract

Background: Obstructive sleep apnea (OSA) is a prevalent and serious medical condition characterized by repeated complete or partial obstructions of the upper airway during sleep and is prevalent in 2% to 4% of working middle-aged adults. Nasal continuous positive airway pressure (CPAP) is the gold-standard treatment for OSA. Because compliance rates with CPAP therapy are disappointingly low, effective interventions are needed to improve CPAP compliance among patients diagnosed with OSA.

Objective: The aim was to determine whether wireless telemonitoring of CPAP compliance and efficacy data, compared to usual clinical care, results in higher CPAP compliance and improved OSA outcomes.

Methods: 45 patients newly diagnosed with OSA were randomized to either telemonitored clinical care or usual clinical care and were followed for their first 2 months of treatment with CPAP therapy. CPAP therapists were not blinded to the participants' treatment group.

Results: 20 participants in each group received the designated intervention. Patients randomized to telemonitored clinical care used CPAP an average of 4.1 ± 1.8 hours per night, while the usual clinical care patients averaged 2.8 ± 2.2 hours per night (P = .07). Telemonitored patients used CPAP on 78% ± 22% of the possible nights, while usual care patients used CPAP on 60% ± 32% of the nights (P = .07). No statistically significant differences between the groups were found on measures of CPAP efficacy, including measures of mask leak and the Apnea-Hypopnea Index. Patients in the telemonitored group rated their likelihood to continue using CPAP significantly higher than the patients in the usual care group. Patients in both groups were highly satisfied with the care they received and rated themselves as "not concerned" that their CPAP data were being wirelessly monitored.

Conclusions: Telemonitoring of CPAP compliance and efficacy data and rapid use of those data by the clinical sleep team to guide the collaborative (ie, patient and provider) management of CPAP treatment is as effective as usual care in improving compliance rates and outcomes in new CPAP users. This study was designed as a pilot—larger, well-powered studies are necessary to fully evaluate the clinical and economic efficacy of telemonitoring for this population.

(J Med Internet Res 2007;9(2):e14) doi:10.2196/jmir.9.2.e14

KEYWORDS

Continuous positive airway pressure therapy; CPAP; sleep apnea syndromes; treatment compliance; telemedicine; randomized controlled trial

Introduction

Sleep-disordered breathing is a generic diagnostic term broadly used to describe apnea (cessation of airflow), hypopnea (reduction in airflow), and other breathing irregularities that occur during sleep. It is a common, albeit underdiagnosed, chronic condition in the adult population, with up to 4% of females and 9% of males experiencing at least 15 episodes of apnea and/or hypopnea per hour of sleep [1-4].The sleep-disordered breathing syndrome known as obstructive sleep apnea/hypopnea (OSA) entails repetitive occlusion or narrowing of the upper airway, which results in intermittent decreases in oxyhemoglobin (blood oxygen desaturation) and increases in blood carbon dioxide levels and, in turn, leads to frequent arousals from sleep. The primary and most measured symptom of OSA is the daytime drowsiness and fatigue that follow the nightly sleep disruption. The seriousness of OSA, however, is underscored in aggregate clinical profile by significant comorbidities, including obesity, depression, ischemic heart disease, stroke, hypertension, and increased risk of automobile-related accidents, and by the extreme outcome of premature death [1,5,6].

Nasal application of continuous positive airway pressure (CPAP) [7] is the treatment of choice for this condition, involving use of both an air flow generator and a mask to supply a constant stream of pressure through the pharyngeal airway during sleep [8]. Despite the documented clinical efficacy of CPAP, many—perhaps most—patients have difficulty adhering to treatment owing to one or more of a variety of discomforting side effects such as pressure intolerance, claustrophobic reaction to the nasal mask, mask dislodgement, and machine noise. Estimates of the noncompliance rates among OSA patients who start a CPAP regimen but terminate treatment within one year are as high as 50% or more [9]. Among patients who persevere to the one year mark or later, most demonstrate only partial compliance by not using CPAP for the entire night as prescribed. Published objective compliance rates from studies conducted in the United States range from 3.3 to 5.3 hours per night [10-12], with one study showing that only 6% of patients used the machine for ≥ 7 hours per night on at least 70% of the nights monitored [10]. As these study findings clearly indicate, effective interventions are needed to improve CPAP compliance among the OSA-diagnosed patient population at large.

Over the past decade, a variety of psychoeducational and technological interventions designed to improve CPAP compliance have been developed and tested. To date, the scope of psychoeducational interventions includes telephone follow-ups with dissemination of OSA-CPAP literature to patients [13], intensive patient education and support protocols [14-16], and motivational enhancement programs [17]. While most of these approaches have had only modest impact on compliance, the intensive support protocols have had greater impact but require a substantial amount of time and resources to implement. Technological advances include more comfortable interfaces, variable positive airway pressure (PAP) levels responsive to the respiratory cycle, and humidification. Although clinical logic and experience suggest that these technological improvements can enhance compliance by increasing patient comfort, a large-scale systematic review of the applications and outcomes reveals that they, on average, result in little, if any, effect on CPAP compliance rates [18]. In the present study, the aim was to achieve better CPAP compliance and clinical efficacy by combining and integrating the most promising elements of both psychoeducational interventions and technological innovations. This design imperative led us to deploy a recently developed telemedicine system that allows tailored management of OSA-CPAP patients through wireless monitoring of "time at prescribed pressure" and transmission of those CPAP compliance- and efficacy-relevant data to care providers in 24-hour cycles.

This pilot intervention was informed and shaped by three major trends in the findings of prior studies, each predicated on a departure from usual care for OSA-CPAP patients. Usual care entails initial patient-specific titration of the CPAP device to the critical pressure needed to keep the upper airway open during sleep, a follow-up telephone call from the CPAP provider or sleep physician's office to check on the patient's comfort and usage within one week of CPAP initiation in the home environment, and subsequent in-office visits starting several weeks after CPAP initiation and continuing thereafter as needed. The first salient trend in prior findings is that patients' self-reported difficulties with CPAP as well as their subjective compliance do not provide sufficient or reliable information to guide appropriate clinical management of OSA under usual care conditions. For example, engaged in a novel therapy, a patient may not be aware of excessive mask leakage during the initial week of CPAP and so will fail to report the problem during the one-week follow-up call. Uncorrected mask leakage, by hampering critical pressure delivery, can have a seriously deleterious impact on compliance. The second trend is that objective compliance and efficacy data are typically not obtained in a timely manner. There are, in general, three standard patient-dependent modes by which a clinical sleep team obtains the objective compliance data recorded as machine-on time by the CPAP unit: (1) the patient brings the CPAP unit into the office and the sleep team physically downloads the data, (2) the patient transmits the data from home via a telephone line, and (3) the patient removes a memory card from the CPAP unit and either mails it in or brings it into the office during a follow-up visit. Much as compliance with medical regimens can be problematic [19], patient transfer of the CPAP data to the clinical team can be equally erratic. The third trend is that compliance patterns are established early in treatment initialization, so time and problem resolution are of the essence in the effort to establish a pattern of treatment compliance [20]. A sleep team's ability to detect problems and intervene early in therapy may improve a patient's early response to CPAP therapy and increase the likelihood that the patient will become a regular and compliant user, thereby enhancing clinical outcome.

Given the very limited number of prior CPAP compliance- and efficacy-related studies within the telemedicine arena, the present study breaks new ground by attempting to counter the known shortcomings of usual care in both telemonitoring design and to answer the overarching question: Does more quickly delivered advice and counsel from providers to patients about developing objectively measured in-home compliance patterns

and problems translate to enhanced treatment compliance levels sufficient for improvements in patient health status and outcomes?

Methods

Design

This was a randomized controlled pilot study approved by the local Institutional Review Board. Newly diagnosed OSA patients who met inclusion criteria were asked to participate. Patients were randomized to either the usual clinical care (UCC) group or the telemonitored clinical care (TCC) group. Both groups of patients received the monitoring device and were followed for an intervention period of 2 months.

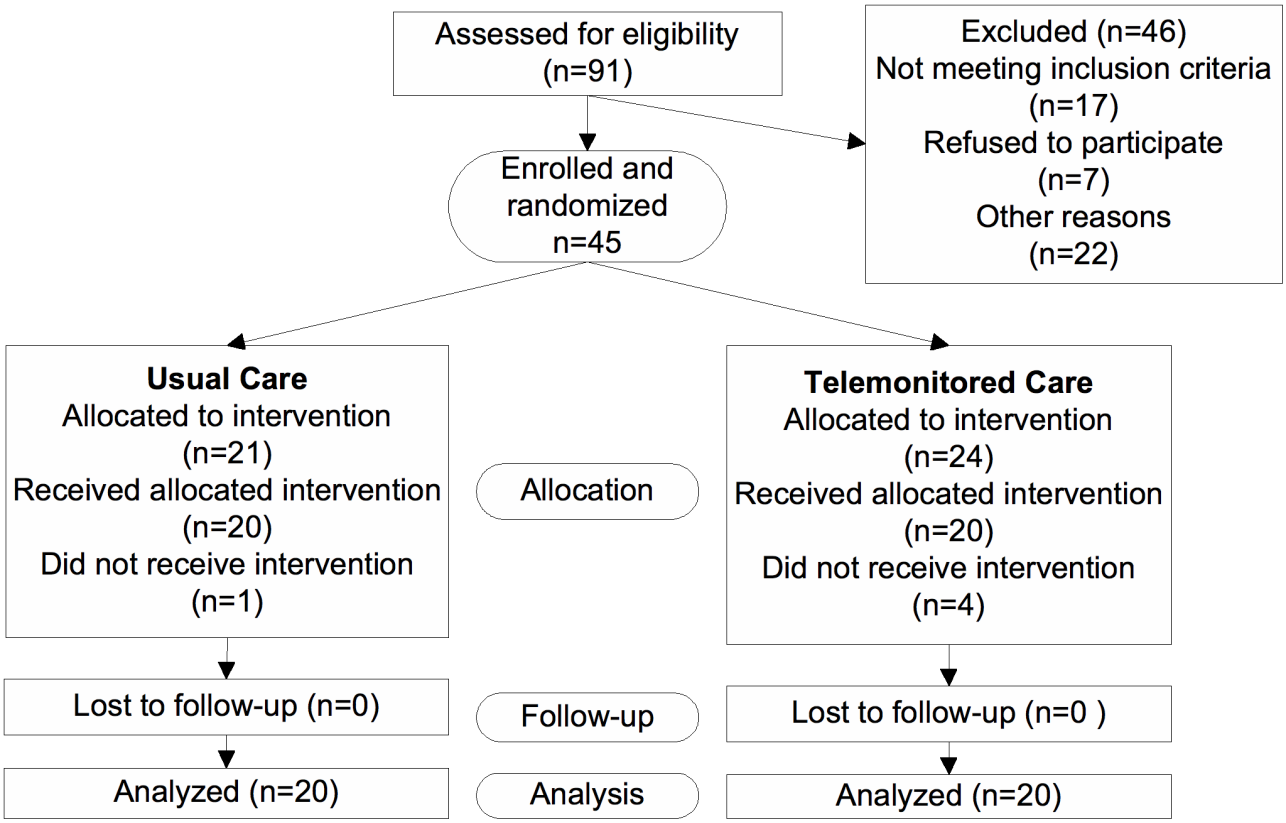
Participants

Participants were patients at the Veterans Affairs San Diego Healthcare System (VASDHS) who were referred to the sleep clinic by their physician for suspicion of OSA. All patients had their sleep recorded with the Stardust sleep recording system (Respironics, Pittsburgh, PA), which monitors heart rate, oximetry, nasal airflow, chest wall movement, and body position. Patients diagnosed with OSA based on the results of their sleep study and clinical history were recruited, consented, and enrolled if they met all of the following criteria: diagnosis of moderate-to-severe OSA, defined as an Apnea-Hypopnea Index (AHI) ≥ 15 events per hour; naive to CPAP therapy; stable sleep environment (operationally defined as a permanent address, requisite for wireless monitoring); and at least 18 years of age. An AHI of ≥15 was chosen in an effort to be consistent with current OSA guidelines and practice parameters [8,9,21,22].

Patients were excluded from the study if they met any one of the following criteria: allergies or sensitivity to the mask or mask material; previous use of any other PAP device (eg, bi-level PAP, auto-adjusting PAP); current use of prescribed supplemental oxygen; or significant comorbid medical conditions that would prevent the patient from completing the protocol. Significant comorbidities were defined as any medical or mental health condition that could interfere with the daily use of CPAP. Additionally, patients were excluded if they lived in a geographically unsuitable region (ie, outside of the wireless network coverage area). A total of 91 patients at the VASDHS Sleep Clinic either signed or gave verbal consent to be contacted so they could learn more about the study. From these 91 patients, 46 were either were not interested in study participation or did not satisfy the inclusion and exclusion criteria.

The remaining 45 patients signed consent forms and were enrolled into the study. The study took place from October 2004 to August 2006. Figure 1 illustrates the patient flow and randomization scheme. The randomization scheme was concealed until the time at which the intervention was assigned. The randomization scheme was generated by the project statistician and carried out by research staff immediately after the informed consent procedure and the completion of the baseline questionnaires. Using random assignment, 21 patients were assigned to the TCC arm and 24 patients to the UCC arm, with 20 participants in each group receiving the designated intervention. There were five CPAP "rejectors," or patients who decided within the first day or two that they did not want to pursue CPAP as the primary treatment for their OSA. Our study did not have a "run-in" period, which could have helped identify these patients prior to the intervention.

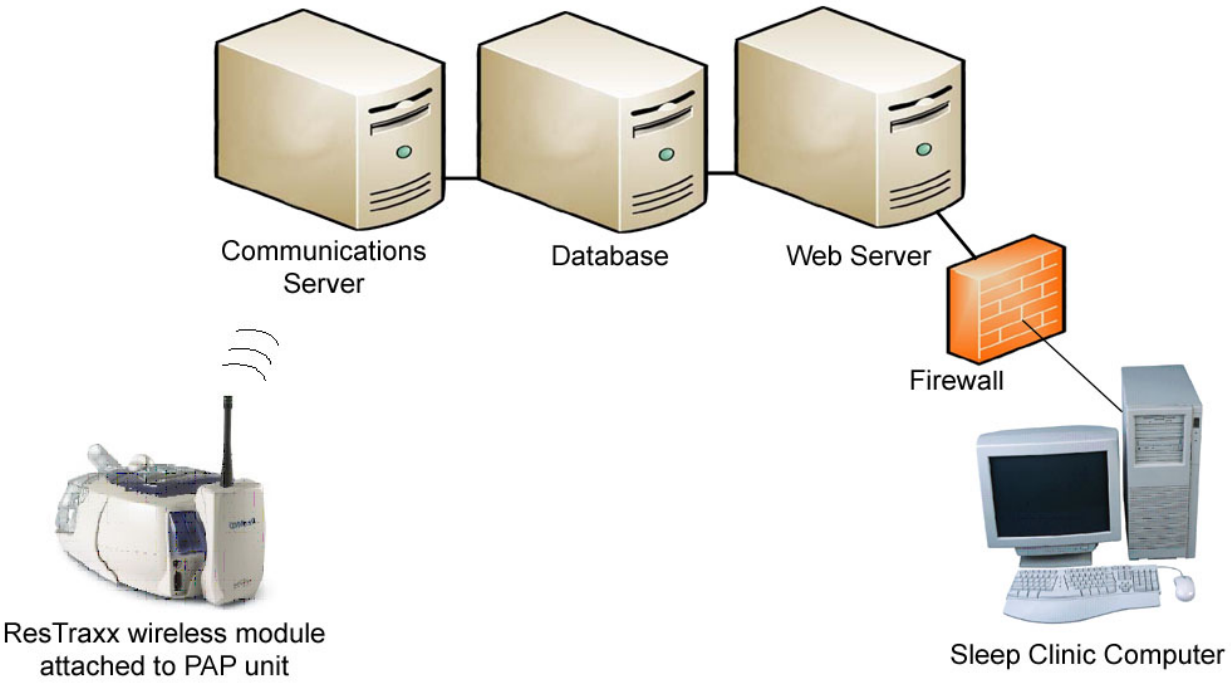
Figure 1. Participant flow and randomization chart



Apparatus

Each participant was provided with an AutoSet Spirit flow generator unit (ResMed Corp, Poway, CA) set to fixed-mode pressure, which was equipped with the HumidAire 2i heated humidifier (ResMed Corp, Poway, CA). Each participant was provided a compatible nasal or full-face mask; nasal pillows were not used in this study.

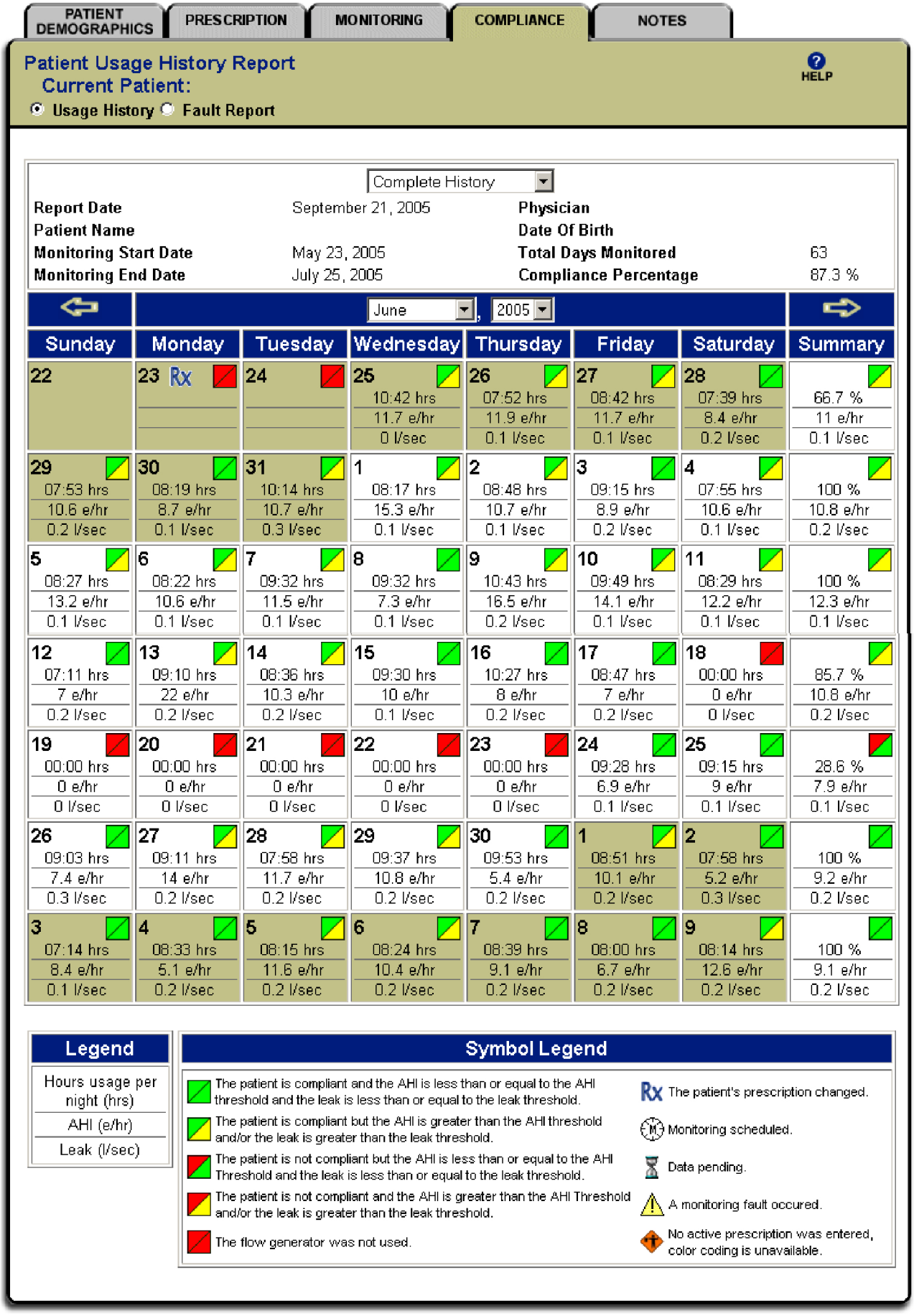
Figure 2. Model of wireless telemonitoring of CPAP compliance and efficacy.



All CPAP flow generators used in the study were outfitted with a ResTraxx wireless transmitter (ResMed Corp, Poway, CA) (Figure 2) to allow the remote data collection of compliance and efficacy information for both groups and to reduce any bias that might be associated with one group (TCC) and not the other (UCC) having the extra attachment. While the patient sleeps, data are collected. The next day, the de-identified data are transmitted to a computer server that meets Health Insurance Portability and Accountability Act (HIPAA) security and privacy standards. The data then become available on a sleep clinic's computer via connection to a password-protected website where clinicians can access the nighttime CPAP data the following afternoon.

Only research and clinical staff had secured access to the data via a standard browser and entry into the ResTraxx Data Center (RDC), the patient and data management website designed for 24/7 access to telemonitored compliance and efficacy data. Per each 24-hour cycle of data transmission, the ResTraxx Data Center website displays the data in a calendar format that reveals daily and weekly data trends. The Multimedia Appendix provides four screenshots of the ResTraxx Data Center website, including patient demographics, prescription, monitoring, and compliance (the latter screen is also shown in Figure 3). Other wireless monitoring systems may be available that allow a similar "passive" procurement of treatment adherence and efficacy data.

Figure 3. Screenshot of the ResTraxx Data Center website, showing the compliance tab with sample compliance and efficacy data for one month of CPAP use for a hypothetical patient. The thresholds specified on another tab allow for color-coding of efficacy and compliance data — this color-coding allows for quick review of how well any one patient is doing on CPAP. Specific data values are also provided, which can aid clinical management. The symbol legend at the bottom of the tab explains the meaning of each symbol used in this tab (see also Table 2).



Groups

All Participants

Regardless of group assignment, all patients who participated in this study had identical CPAP setups. The initial set-up visit consisted of sleep apnea education, sleep study review, and CPAP instruction and mask-fitting. During this visit, a study staff member educated the patient about sleep apnea and explained the patient's baseline sleep study results. Each patient was then instructed on how to use CPAP and was fitted for a mask. The patient wore the mask while the CPAP unit was set at the prescribed pressure for approximately 10 minutes for mask adjustment and assessment of pressure tolerance. Both groups were shown how to attach the ResTraxx wireless device to their CPAP units and were provided with written instructions on CPAP use.

Usual Clinical Care (UCC) Group

Patients randomized to UCC were treated according to the prevailing standard of care for OSA patients at the VASDHS CPAP Clinic. Usual care consisted of a 1-week telephone call after CPAP initiation and a 1-month in-office follow-up visit by CPAP clinic staff. Patients were encouraged to call the clinic any time they had a problem or concern. CPAP compliance and efficacy data were downloaded at the 1-month time point to help direct clinical management.

Telemonitored Clinical Care (TCC) Group

The essence of the TCC intervention is the ability to telemonitor compliance and efficacy data for each patient on a daily basis from the first day of treatment and to act on those data collaboratively, and in partnership, with the patient. Collaborative management refers to the joint decision making and partnership between provider and patient and is characterized by communication, negotiation, and consideration of important patient factors and preferences. Patients in this group had their objective flow generator data monitored as frequently as needed per specified clinical pathways throughout the active 2-month treatment period. The frequency and nature of the clinical interactions depended on both the objectively measured nightly data values and subjective patient reports.

TCC Parameters and Thresholds

Telemonitoring included review of both compliance and efficacy data. Compliance data encompassed details on how many total hours the PAP unit was used each night at therapeutic pressure. Efficacy data consisted of the amount of mask leakeage (L/s) and the AHI (total number of apneas and hypopneas per hour of sleep). Thresholds for each parameter were set by the study team using the password-secured interactive ResTraxx Data Center website (Table 1).

Table 1. Specifications of thresholds for each parameter

Parameter	Threshold
CPAP compliance	4 hours/night
Apnea-Hypopnea Index	10 events/hour of sleep
Mask leak	0.4 L/s

Website Color-Coding Scheme

The ResTraxx Data Center calendar display (see also Figure 3) summarized the compliance (upper-left diagonal) and AHI / mask leak measures (lower-right diagonal) for each night in a color-coded box-shaped icon. Table 2 presents the color-code scheme of these icons and provides a data description, an interpretation in relation to the threshold value for each parameter (see Table 1), and a recommended general course of action per each actual measure. While each night's measures were color-coded to ease visual identification of out-of-range values when tracking large numbers of patients, the display for each day contained the specific data values for compliance and efficacy as well. In addition, each week's display included a summary box providing mean values for each of the parameters for that week.

Table 2. Color-coding scheme summarizing compliance and efficacy measures on the ResTraxx Data Center website

Code	Description			Interpretation	Action
green/... 	Compliance ≥ 4 hours/night			Compliance at or above threshold	Compliance within limits, no intervention necessary
red/... 	Compliance < 4 hours/night			Compliance below threshold	Consider intervening to increase compliance
.../green 	AHI < 10	and	Leak < 0.4 L/s	Leak and AHI below threshold	Efficacy within limits, no intervention necessary
.../yellow 	AHI < 10	and	Leak ≥ 0.4 L/s	Either leak or AHI above threshold	Identify which is high and intervene as necessary
		or			
	AHI ≥ 10	and	Leak < 0.4 L/s		

Note: The only color code not shown is red/red, which indicates that CPAP was not used on the particular night of monitoring.

Clinical Pathways

For the purposes of this study, we defined clinical pathways for the interventionists to follow. The pathways specified how frequently the clinical team would check the CPAP compliance and efficacy data values on the ResTraxx Data Center website for each patient. Human monitoring is designed to be more intensive in the earlier stages of treatment, with a gradual tapering off over time if patterns of CPAP compliance are established. A green/green pathway (ie, when all three parameters of compliance, AHI, and mask leak are within normal limits) specified this gradually attenuated monitoring schedule. A red/yellow pathway specified the general course of action required when one or more parameters were not within the normal range of values. In each case, the clinician and the patient collaboratively assessed the source of the problem, considered alternative solutions, and selected a corrective measure. The clinical team then continued close monitoring until each parameter was back within normal range. The red/yellow pathway referred the clinician to an intervention matrix of corrective actions to consider for each problem (Table 3).

Table 3. CPAP clinician intervention management matrix for TCC group (adapted from [23])

	Symptoms	Cause	Correction
Mask	Dry eyes	Leak	Verify mask fit/size
	Irritated skin	Head gear too tight	Readjust head gear
	Pressure sores	Incorrect mask fit	
Nasal	Dry nose, mouth, throat	Lack of humidified air	Use heated humidifier
	Nasal congestion	Mouth breathing	Try chin strap
Pressure	Feeling of need for more or less air	Incorrect pressure (too low or high)	Verify prescribed pressure
	Chest discomfort	Claustrophobia	Add short periods of use during the day
Miscellaneous	CPAP noise	Blocked air intake	Check air filter
	Bed partner disturbance	Multiple factors	Move unit further from bed
	Still snoring and/or sleepy	Pressure too low	Consider pressure increase

Data Collection

Questionnaire data were collected at baseline and postintervention. The Measures section below describes each assessment tool and provides operational definitions. Flow generator data from all patients were both wirelessly transmitted to the secure ResTraxx Data Center server and manually downloaded by research staff at the end of the 2-month monitoring period. Data obtained wirelessly were compared to data obtained via manual download to verify accuracy in transmission and confirm the 100% accuracy rate reported previously in the literature for wireless transmission of CPAP data [24]. The study period was composed of 2270 nights for wireless data transmission. However, timely transmission failures occurred on 131 of those nights, 5.7% of the total, owing mainly to minor network issues related to residence location with intermittent coverage. Over the 2088 nights of data collection available for comparison, there were no discrepancies

between the two data sources, that is, 100% accuracy was achieved in this study between data obtained wirelessly and data downloaded manually.

Measures

This study assessed a number of measures both at baseline and postintervention from a number of domains, including sleep study data, CPAP-related data, OSA symptoms, health-related quality of life, and psychological factors (ie, depressive symptoms). Sleep study data were obtained from the baseline Stardust sleep recording system measures that established the diagnosis of OSA and included the AHI. Sleep apnea symptom scales were based on previously published scales of sleep apnea symptom frequency that are reliable, valid, and highly predictive of OSA [25,26]. The Epworth Sleepiness Scale (ESS) [27,28] is a widely used subjective measure of excessive daytime sleepiness that queries likelihood of falling asleep in eight different situations on a scale from 0 (not likely) to 3 (highly likely). Scores on the ESS range from 0 to 24, with higher scores reflective of greater daytime sleepiness. In addition, the research team used a visual analog scale to measure sleepiness. The Center for Epidemiological Studies Depression Scale (CES-D) was used to assess depressive symptomatology [29,30].

A fundamental methodological advantage in studying CPAP compliance is that compliance is measured objectively via a device-internal clock counter. The primary measure used in this study was the number of hours per night that the unit was used at the prescribed pressure. For CPAP device efficacy, three facets of PAP efficacy were measured by the AutoSet Spirit flow generator: mask leak (L/s), pressure (cm H_20), and AHI. AHI measurements by AutoSet Spirit have been shown to be highly correlated with the measures recorded by polysomnography [31]. Identical devices were provided to both the TCC and UCC groups to eliminate any potential variation attributable to the use of different devices.

OSA-specific health-related quality of life was assessed using the Functional Outcomes of Sleep Questionnaire (FOSQ). This is a 32-item self-report measure that assesses the impact of disorders of excessive sleepiness on multiple activities of daily living [32].

Clinician satisfaction and patient satisfaction were assessed by questionnaires that include both Likert-type items and open-ended questions. Participants were asked to rate their overall satisfaction with care (1 = poor; 5 = excellent), their likelihood of continuing to use CPAP (1 = not likely; 5 = highly likely), and their concern about being monitored wirelessly (1 = not concerned; 5 = highly concerned).

CPAP self-efficacy refers to OSA patients' confidence or belief that they can adhere to the regimen necessary to manage their OSA with CPAP. The CPAP self-efficacy scale is a 5-item self-report scale that was developed and validated by the primary author [12].

Communication with the health care team was assessed by a 3-item self-report measure that measures the frequency with which patients use certain communication behaviors, including preparing a list of questions, asking questions about things they want to know and things they don't understand about their treatment, and discussing any personal problems that may be related to their illness [33]. Respondents are asked to answer using a 6-point Likert-type response ranging from never (0) to always (5).

Statistical Analysis

For comparison between usual and telemonitored groups at postintervention, an unpaired *t* test was used for normally distributed variables, and the Mann-Whitney test was used for non-normally distributed variables. The improvement of variables from baseline was tested by paired *t* test.

Power Analysis

We based the power analysis on the effect size from one study that compared intensive clinical support with standard clinical support [15]. The CPAP compliance level in the intensive support group (5.4 ± 1.9 hours/night) was higher than in the standard care group (3.9 ± 2.5 hours/night) over the first month of treatment. In terms of effect sizes, this approximates a d-index of .65. Based on this effect size, an alpha level of .05, and a single-sided hypothesis test, 42 patients would be required to have a power level of .70, which would be an acceptable power level for this pilot study.

The one variable that is difficult to accurately estimate prospectively is the standard deviation (SD) for the treatment compliance; to the extent that SD is high relative to the mean difference, power is lower.

Results

The participants who completed this study were primarily male (98%), older, and overweight and had moderate-to-severe OSA. Table 4 provides details on the baseline characteristics of the sample and shows that the two groups did not differ on any of the measured scales at baseline. The sample was primarily middle-aged and overweight. In addition, approximately 61% were Caucasian, 15% Black, 10% Filipino, 7% Hispanic, 2% Asian or Pacific Islander, and 5% (self-described as) "other." All of the participants reported education levels at or above the high school graduate level, with 61% having the equivalent of 14 years of education; 50% of the sample was married; 42% was retired, 37% was working (full or part time), and the remaining 21% was either unemployed, unable to work, or students. The mean (SD) for the education variable was 14.7 years (3.2) and 14.3 years (2.7) for the UCC and TCC groups, respectively. The difference between these means was not statistically different (P = .62).

Table 4. Baseline patient characteristics and treatment period, by study arm

	Total Group		TCC		UCC		
Characteristic	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	*P* value*
Age (years)	59 ± 14.3	23-80	60 ± 10.8	42-80	58 ± 13.7	23-74	.50
Body mass index (kg/m^2)	32.8 ± 5.7	26-45	33.3 ± 4.9	25-46	30.5 ± 5.1	25-42	.22
Apnea-Hypopnea Index	39 ± 16.8	21-94.7	44.8 ± 17.9	23.3-93.7	37.6 ± 14.3	21-62.5	.29
CPAP pressure (cm H_20)	10.3 ± 1.6	8-13	10.4 ± 1.3	8-13	9.7 ± 1.4	8-13	.29
Sleep apnea symptoms: night	3.1 ± 0.7	1.8-4.7	3.2 ± 0.8	1.8-4.7	2.9 ± 0.6	1.9-3.9	.30
Sleep apnea symptoms: day	2.6 ± 0.8	0.9-4.7	2.6 ± 0.9	0.9-4.7	2.7 ± 0.8	1.4-4.6	.75
Epworth Sleepiness Scale	12.6 ± 5.8	4-23	13.7 ± 5.8	4-23	11.7 ± 4.4	2-20	.25
Sleepiness visual analog scale	5.8 ± 2.6	0-10	5.6 ± 3.0	0-10	6.1 ± 2.3	2-10	.77
Functional Outcomes of Sleep Questionnaire	13.8 ± 3.8	6.2-19.3	13 ± 4.5	5.3-19.3	14.2 ± 3.4	3.9-20	.54
Treatment period (number days monitored)	59.6 ± 4.0	43-63	60 ± 3.0	51-63	60.2 ± 2.8	55-63	.26

**t* test comparing TCC vs UCC

CPAP Compliance

Participants were followed for an average treatment period of 60 ± 4.0 days. The primary outcome measure of the study was level of CPAP treatment compliance: the TCC group used CPAP 4.1 ± 1.8 hours per night, while the UCC control group used CPAP 2.8 ± 2.2 hours per night (*P* = .07). This difference represented an effect size of 0.65, with membership in the telemonitored group resulting in a 46% increase in CPAP use relative to usual care. There was also a trend for the groups to differ on the percentage of nights with any amount of CPAP use (TCC: 78% ± 22%; UCC: 60% ± 32%; *P* = .07). Table 5 provides details on other measures of CPAP compliance and shows that while none reached statistical significance between study arms, all were in the direction of greater CPAP use in the TCC group.

Table 5. CPAP compliance and efficacy data summary, postintervention

	Total Group		TCC		UCC		
Characteristic	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	*P* value*
Compliance							
Use (hours/night) (all days)	3.5 ± 2.1	0.2-6.8	4.1 ± 1.8	0.12-6.8	2.8 ± 2.2	0.2-6.2	.07
Use (hours/night) (days used)	4.4 ± 2.2	0.4-8.7	5.0 ± 1.8	0.5-7.7	3.8 ± 2.3	0.4-7.8	.10
% nights with CPAP use > 0 hours of use	65 ± 31	0-100	78 ± 22	24-98	60 ± 32	5-100	.07
% nights with CPAP use > 4 hours of use	44 ± 32	0-93	52 ± 27	0-93	37 ± 34	0-89	.16
Efficacy							
Apnea-Hypopnea Index (events/hour)	6.8 ± 5.3	0.1-25.6	7.9 ± 4.1	3.4-19.1	5.0 ± 4.0	0.1-13	.04
Arousal Index (events/hour)	1.5 ± 2.2	0.1-11.6	1.4 ± 1.3	0.2-4.9	1.2 ± 1.4	0.02-4.0	.64
Apnea-Hypopnea Index change	35.6 ± 16.4	12.5-89.3	38.1 ± 18.4	18.6-89.3	32.2 ± 14.8	12.5-58.8	.32
Mask leak, median (L/s)	0.18 ± .36	0-2.1	0.12 ± 0.11	0.01-0.44	0.26 ± 0.51	0-2.1	.29
Mask leak, 95th percentile (L/s)	0.44 ± .36	0.1-2.2	0.38 ± 0.18	0.17-0.82	0.50 ± 0.47	0.1-2.2	.31
Mask leak, maximum (L/s)	0.71 ± .41	0.1-2.2	0.68 ± 0.25	0.26-1.4	0.75 ± 0.54	0.1-2.2	.62

**t* test comparing TCC vs UCC

CPAP Efficacy

The two classes of CPAP efficacy measures were effect of CPAP on disease severity (AHI) and effect of CPAP on reduction of mask leak. The mean AHI across the 2-month intervention period was significantly different between the groups (TCC: 7.9 ± 4.1; UCC: 5.0 ± 4.0; $P = .04$), with the TCC group having higher AHI values. However, a more appropriate analysis than direct mean comparison is to take into account the baseline AHI by calculating the reduction in AHI from baseline. Therefore, change in AHI was calculated as baseline AHI (as measured by diagnostic overnight sleep study) minus the mean AHI (as measured by the flow generator device over the course of the intervention period). Table 5 shows the results of the *t* test for change in AHI; there was no statistically significant difference between the groups (TCC: 38.1 ± 18.4; UCC: 32.2 ± 14.8; $P = .32$).

With respect to the effect of CPAP on mask leakage, no statistically significant differences in mask leak level were found between the groups on three difference measures: median leak, 95th percentile leak, and maximum leak. See Table 5 for more details.

Outcomes

No statistically significant differences were found for any of the outcome measures, as summarized in Table 6. The TCC group had slightly higher scores on the communication with health care team measure than the UCC group (3.3 vs 2.5; $P = .07$).

Table 6. Outcome measures, postintervention

	Total Group		TCC		UCC		
Characteristic	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	*P* value*
Sleep apnea symptoms: night	2.4 ± 0.62	1.4-4.0	2.4 ± 0.66	1.4-3.7	2.3 ± 0.61	1.4-4.0	.96
Sleep apnea symptoms: day	2.1 ± 0.89	0.6-4.1	2.0 ± 1.1	0.57-4.1	2.1 ± 0.65	0.92-3.21	.57
Epworth Sleepiness Scale	9.6 ± 5.9	2.0-21.0	9.2 ± 6.6	2.0-21.0	9.9 ± 5.2	3.0-20.0	.72
Sleepiness visual analog scale	4.3 ± 2.9	0.0-10.0	3.8 ± 3.4	0.0-10.0	4.8 ± 2.2	2.0-9.0	.35
Functional Outcomes of Sleep Questionnaire	14.8 ± 4.6	1.8-19.8	15.2 ± 5.0	1.8-19.8	14.4 ± 4.2	1.8-19.9	.61
Center for Epidemiological Studies Depression Scale	8.5 ± 6.4	0.0-23.0	8.6 ± 7.0	0.0-23.0	8.3 ± 5.8	0.0-19.0	.91
CPAP self-efficacy	3.8 ± 1.1	1.0-5.0	4.0 ±1.2	1.0-5.0	3.5 ± 0.83	2.0-5.0	.21
Communication with health care team	2.9 ± 1.2	0.7-5.0	3.3 ± 1.2	0.7-5.0	2.5 ± 1.2	0.7-5.0	.07

**t* test comparing TCC vs UCC

Patient Satisfaction

Patients in the TCC group rated their likelihood to continue using CPAP significantly higher than the UCC group (4.8 vs 4.3; $P = .05$). The TCC and UCC groups did not differ in their ratings of satisfaction with care (both groups highly satisfied) or their concerns about being wirelessly monitored (both groups not concerned).

Discussion

The results of this pilot study suggest that use of telemonitored CPAP compliance and efficacy data to guide the collaborative clinical management of CPAP treatment appears to be as good as usual care in its effect on compliance rates and outcomes in new CPAP users. Given that CPAP use patterns are established early in the treatment initialization process, the monitoring of compliance and efficacy data early in this process, updated in continuous 24-hour periods, to oversee and, if necessary, intervene in treatment is one method clinicians have available to them.

While the usual care compliance rate of 2.8 hours per night is low compared to other published CPAP compliance intervention studies [14-16], it is within close range of compliance levels reported in a comparable studies of the veteran population [34,35], which is on the order of 3.1 hours per night. The TCC group in the present study had a compliance rate of 4.1 hours per night after the 2-month period, which represents a 46% increase in compliance over the mean compliance level of the UCC group (2.8 hours per night) and is equivalent to an effect size of 0.65. According to Cohen's classification of effect sizes, this would represent a large effect [36].

This study did not find a statistically significant effect at the .05 alpha level for the TCC group on the primary outcome measure of CPAP compliance. The initial power analysis, described above in the Methods section, underestimated the residual variance. With the actual variance, a sample of the size that we collected had a low power, implying that a "failure to reject" at the .05 alpha level results in an unacceptably high type two error rate. The obtained effect size would be clinically significant if it could be established to be reliable, and the only way to do that is to perform a study with a larger sample, as we propose.

The present study confirmed the 100% accuracy rate of wireless data transmission in comparison to manually downloaded data. There was some concern at study outset over the potential for data loss via wireless transmission; however, we found the

loss was negligible. While some data may not have come through on the website (either due to residence location with intermittent coverage or to initial patient problems in attaching the units), the data were always available on the flow generator unit itself. It should be noted that once the wireless unit was properly attached, data from previous nights that are stored on the flow generator device can be re-transmitted and obtained wirelessly. As a further safeguard, data can always be manually downloaded directly from the flow generator device during a patient visit. Our experience was that the data could be obtained within a few days and did not negatively impact clinical management. Future generations of flow generator devices may benefit from the pre-installation of an internal wireless device, thereby avoiding potential patient problems with attaching the wireless unit, especially for older adult patients with less technical experience or limited manual dexterity.

The study design dictated that both groups have ResTraxx wireless devices attached to their flow generator units. Prior research has suggested that the mere use of medical devices or components can result in a placebo effect [37]. As such, the UCC group did not necessarily receive usual and customary care because of the presence of the wireless unit. The main implication of this difference is that this study may have underestimated the effect of TCC on CPAP compliance and efficacy measures compared to true usual care. To control and better assess this effect, a third study arm composed of UCC with no wireless device may be added in our future studies of TCC.

Mask leak was not significantly different between the groups. This result is likely owing to the fact that both groups had median mask leak levels that were within normal limits. Likewise, the AHI measured by the CPAP unit across the intervention period was not significantly different between the two groups. This may be a function of measurement (though there is evidence that the AHI measured by the flow generator device used in this study is quite comparable to the AHI measured by polysomnography [31]), sample size, or even limited recording time. It should be noted that limited recording time can play a role because the number of apneas and hypopneas are only measured during the time periods when the CPAP is used. In the present study, CPAP on-time was lower in the UCC group, so their AHI data may be less reliable.

TCC appeared to have no effect on outcomes relative to UCC. Only communication with the health care team was close to being statistically significant at the .05 alpha level. This measure assesses the frequency with which patients use certain communication behaviors, including preparing a list of questions, asking questions about things they want to know and things they don't understand about their treatment, and discussing any personal problems that may be related to their illness. Focusing on the content and process of patient-sleep provider interactions in future CPAP compliance studies may prove to be a fruitful area of research, particularly in looking at the differences between face-to-face interactions compared to telephone calls.

It can be difficult to reliably measure outcomes in OSA patients given the heterogeneity of the clinical presentation. For example, it is well documented that measures of sleepiness and the current gold-standard measure of disease severity, the AHI, correlate only modestly in the 0.30 range [38]. This is largely due to the fact that OSA is a syndrome, so better characterization of subsets of OSA patients (ie, patient stratification by syndrome composition and severity) would allow more precise outcome measurement. Stratified sampling could not be conducted in the present pilot study owing to the small sample size. We did, however, conduct secondary analyses of AHI groups (moderate: AHI 15-29.99; severe: AHI ≥ 30), but all of the reported findings held within each group.

The present study is most comparable to three prior CPAP telemedicine studies. One study utilized a daily computer-based telephone system to monitor patients' self-reported compliance behavior and provide automated counseling through a structured dialogue [39]. The impact of the intervention was not significant in comparison to usual care; however, the findings suggest that concurrent education and reinforcement during the initial and early treatment period are effective countermeasures to patient-reported attenuated compliance. Another study utilized computers to provide daily Internet-based informational support and feedback for problems experienced with CPAP use, again without objective measures of compliance, but found no significant differences between the telemedicine intervention and usual care group at 30 days in patient functional status and satisfaction with CPAP [40]. A third study pilot-tested a teleconferencing intervention in which a nurse visually assessed mask fit and patients' CPAP procedures and provided counseling and reinforcement to a small sample of patients who were trying CPAP again after an initial 3-month period of poor compliance [41]. Although the patient education materials supplied during the initial period did not impact compliance rates, the nurse teleconferencing sessions during the second trial period significantly improved the compliance of the intervention group, suggesting that intensity of one-on-one counseling and feedback by a care provider is a salient variable.

This study was designed as a pilot exploration of the telemedicine intervention in comparison to the standard of care for OSA. Clearly, larger well-powered studies are necessary to follow up on the trends found in the data and to evaluate fully the effect of telemonitoring on compliance, efficacy, and key outcomes in this population.

Acknowledgments

This work was supported in part by ResMed Corp (Poway, CA), VA MREP #02-266, VA IIR #02-275, the Research Service of the VA San Diego Healthcare System, and the Veterans Medical Research Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix

Screenshots of the ResTraxx Data Center website [PDF file (Adobe Acrobat), 216 KB - jmir_v9i2e14_app1.pdf]

References

1. Durán J, Esnaola S, Rubio R, Iztueta A. Obstructive sleep apnea-hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 yr. Am J Respir Crit Care Med 2001 Mar;163(3 Pt 1):685-689 [FREE Full text] [Medline: 11254524]
2. Gibson GJ. Obstructive sleep apnoea syndrome: underestimated and undertreated. Br Med Bull 2004;72(1):49-65. [Medline: 15798032] [doi: 10.1093/bmb/ldh044]
3. Marshall NS, Barnes M, Travier N, Campbell AJ, Pierce RJ, Mcevoy RD, et al. Continuous positive airway pressure reduces daytime sleepiness in mild to moderate obstructive sleep apnoea: a meta-analysis. Thorax 2006 May;61(5):430-434 [FREE Full text] [Medline: 16467072] [doi: 10.1136/thx.2005.050583]
4. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 1993 Apr 29;328(17):1230-1235 [FREE Full text] [Medline: 8464434] [doi: 10.1056/NEJM199304293281704]
5. Nieto FJ, Young TB, Lind BK, Shahar E, Samet JM, Redline S, et al. Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. Sleep Heart Health Study. JAMA 2000 Apr 12;283(14):1829-1836 [FREE Full text] [Medline: 10770144] [doi: 10.1001/jama.283.14.1829]
6. Shepard JW. Hypertension, cardiac arrhythmias, myocardial infarction, and stroke in relation to obstructive sleep apnea. Clin Chest Med 1992 Sep;13(3):437-458. [Medline: 1521412]
7. Sullivan CE, Issa FG, Berthon-Jones M, Eves L. Reversal of obstructive sleep apnoea by continuous positive airway pressure applied through the nares. Lancet 1981 Apr 18;1(8225):862-865. [Medline: 6112294] [doi: 10.1016/S0140-6736(81)92140-1]
8. Kushida CA, Littner MR, Hirshkowitz M, Morgenthaler TI, Alessi CA, Bailey D, et al; American Academy of Sleep Medicine. Practice parameters for the use of continuous and bilevel positive airway pressure devices to treat adult patients with sleep-related breathing disorders. Sleep 2006 Mar 1;29(3):375-380. [Medline: 16553024]
9. Loube DI, Gay PC, Strohl KP, Pack AI, White DP, Collop NA. Indications for positive airway pressure treatment of adult obstructive sleep apnea patients: a consensus statement. Chest 1999 Mar;115(3):863-866 [FREE Full text] [Medline: 10084504] [doi: 10.1378/chest.115.3.863]
10. Kribbs NB, Pack AI, Kline LR, Smith PL, Schwartz AR, Schubert NM, et al. Objective measurement of patterns of nasal CPAP use by patients with obstructive sleep apnea. Am Rev Respir Dis 1993 Apr;147(4):887-895. [Medline: 8466125]
11. Reeves-Hoche MK, Meck R, Zwillich CW. Nasal CPAP: an objective evaluation of patient compliance. Am J Respir Crit Care Med 1994 Jan;149(1):149-154. [Medline: 8111574]
12. Stepnowsky CJ, Marler MR, Ancoli-Israel S. Determinants of nasal CPAP compliance. Sleep Med 2002 May;3(3):239-247. [Medline: 14592213] [doi: 10.1016/S1389-9457(01)00162-9]
13. Chervin RD, Theut S, Bassetti C, Aldrich MS. Compliance with nasal CPAP can be improved by simple interventions. Sleep 1997 Apr;20(4):284-289. [Medline: 9231954]
14. Likar LL, Panciera TM, Erickson AD, Rounds S. Group education sessions and compliance with nasal CPAP therapy. Chest 1997 May;111(5):1273-1277 [FREE Full text] [Medline: 9149582]
15. Hoy CJ, Vennelle M, Kingshott RN, Engleman HM, Douglas NJ. Can intensive support improve continuous positive airway pressure use in patients with the sleep apnea/hypopnea syndrome? Am J Respir Crit Care Med 1999 Apr;159(4 Pt 1):1096-1100 [FREE Full text] [Medline: 10194151]
16. Hui DS, Chan JK, Choy DK, Ko FW, Li TS, Leung RC, et al. Effects of augmented continuous positive airway pressure education and support on compliance and outcome in a Chinese population. Chest 2000 May;117(5):1410-1416 [FREE Full text] [Medline: 10807830] [doi: 10.1378/chest.117.5.1410]
17. Aloia MS, Arnedt JT, Riggs RL, Hecht J, Borrelli B. Clinical management of poor adherence to CPAP: motivational enhancement. Behav Sleep Med 2004;2(4):205-222. [Medline: 15600056] [doi: 10.1207/s15402010bsm0204_3]
18. Haniffa M, Lasserson TJ, Smith I. Interventions to improve compliance with continuous positive airway pressure for obstructive sleep apnoea. Cochrane Database Syst Rev 2004(4):CD003531. [Medline: 15495057] [doi: 10.1002/14651858.CD003531.pub2]
19. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care 2004 Mar;42(3):200-209. [Medline: 15076819] [doi: 10.1097/01.mlr.0000114908.90348.f9]
20. Weaver TE, Kribbs NB, Pack AI, Kline LR, Chugh DK, Maislin G, et al. Night-to-night variability in CPAP use over the first three months of treatment. Sleep 1997 Apr;20(4):278-283. [Medline: 9231953]

21. Kushida CA, Littner MR, Morgenthaler T, Alessi CA, Bailey D, Coleman J, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 2005 Apr 1;28(4):499-521. [Medline: 16171294]
22. Chesson AL, Ferber RA, Fry JM, Grigg-Damberger M, Hartse KM, Hurwitz TD, et al. The indications for polysomnography and related procedures. Sleep 1997 Jun;20(6):423-487. [Medline: 9302726]
23. Zozula R, Rosen R. Compliance with continuous positive airway pressure therapy: assessing and improving treatment outcomes. Curr Opin Pulm Med 2001 Nov;7(6):391-398. [Medline: 11706314] [doi: 10.1097/00063198-200111000-00005]
24. Lankford DA. Wireless CPAP patient monitoring: accuracy study. Telemed J E Health 2004 Dec;10(2):162-169. [Medline: 15319046] [doi: 10.1089/tmj.2004.10.162]
25. Maislin G, Pack AI, Kribbs NB, Smith PL, Schwartz AR, Kline LR, et al. A survey screen for prediction of apnea. Sleep 1995 Apr;18(3):158-166. [Medline: 7610311]
26. Gurubhagavatula I, Maislin G, Pack AI. An algorithm to stratify sleep apnea risk in a sleep disorders clinic population. Am J Respir Crit Care Med 2001 Nov 15;164(10 Pt 1):1904-1909 [FREE Full text] [Medline: 11734444]
27. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991 Dec;14(6):540-545. [Medline: 1798888]
28. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep 1992 Aug;15(4):376-381. [Medline: 1519015]
29. 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 Mar;10(2):77-84. [Medline: 8037935]
30. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Measure 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
31. Gugger M. Comparison of ResMed AutoSet (version 3.03) with polysomnography in the diagnosis of the sleep apnoea/hypopnoea syndrome. Eur Respir J 1997 Mar;10(3):587-591 [FREE Full text] [Medline: 9072989]
32. Weaver TE, Laizner AM, Evans LK, Maislin G, Chugh DK, Lyon K, et al. An instrument to measure functional status outcomes for disorders of excessive sleepiness. Sleep 1997 Oct;20(10):835-843. [Medline: 9415942]
33. Lorig K, Stewart A, Ritter P, Gonzalez V, Laurent D, Lynch J. Outcome Measures for Health Education and other Health Care Interventions. Thousand Oaks, CA: Sage Publications; 1996.
34. Stepnowsky CJ, Marler MR, Palau J, Annette Brooks J. Social-cognitive correlates of CPAP adherence in experienced users. Sleep Med 2006 Jun;7(4):350-359. [Medline: 16713349] [doi: 10.1016/j.sleep.2005.11.004]
35. Stepnowsky C, Palau JJ, Marler MR, Gifford AL. CPAP adherence rates in a large clinical population. Sleep 2006;29(Abstract Supplement):A173.
36. Cohen J. Statistical Power for the Behavioral Sciences. 2nd edition. Hillsdale, NJ: Erlbaum; 1988.
37. Kaptchuk TJ, Goldman P, Stone DA, Stason WB. Do medical devices have enhanced placebo effects? J Clin Epidemiol 2000 Aug;53(8):786-792. [Medline: 10942860] [doi: 10.1016/S0895-4356(00)00206-7]
38. Stradling JR, Davies RJO. Sleep. 1: Obstructive sleep apnoea/hypopnoea syndrome: definitions, epidemiology, and natural history. Thorax 2004 Jan;59(1):73-78 [FREE Full text] [Medline: 14694254] [doi: 10.1136/thx.2003.007161]
39. Demolles DA, Sparrow D, Gottlieb DJ, Friedman R. A pilot trial of a telecommunications system in sleep apnea management. Med Care 2004 Aug;42(8):764-769. [Medline: 15258478] [doi: 10.1097/01.mlr.0000132353.99209.fe]
40. Taylor Y, Eliasson A, Andrada T, Kristo D, Howard R. The role of telemedicine in CPAP compliance for patients with obstructive sleep apnea syndrome. Sleep Breath 2006 Sep;10(3):132-138. [Medline: 16565867] [doi: 10.1007/s11325-006-0059-9]
41. Smith CE, Dauz ER, Clements F, Puno FN, Cook D, Doolittle G, et al. Telehealth services to improve nonadherence: A placebo-controlled study. Telemed J E Health 2006 Jun;12(3):289-296. [Medline: 16796496] [doi: 10.1089/tmj.2006.12.289]

Abbreviations

AHI: Apnea-Hypopnea Index
CPAP: continuous positive airway pressure
OSA: obstructive sleep apnea
PAP: positive airway pressure
TCC: telemonitored clinical care
UCC: usual clinical care
VASDHS: Veterans Affairs San Diego Healthcare System

submitted 13.10.06; peer-reviewed by C Smith, M Aloia; comments to author 03.11.06; revised version received 26.04.07; accepted 08.05.07; published 17.05.07.

Please cite as:
Stepnowsky CJ Jr., Palau JJ, Marler MR, Gifford AL
Pilot Randomized Trial of the Effect of Wireless Telemonitoring on Compliance and Treatment Efficacy in Obstructive Sleep Apnea
J Med Internet Res 2007;9(2):e14
URL: http://www.jmir.org/2007/2/e14/
doi:10.2196/jmir.9.2.e14
PMID:17513285

Original Paper

Following Up Nonrespondents to an Online Weight Management Intervention: Randomized Trial Comparing Mail versus Telephone

Mick P Couper[1], PhD; Andy Peytchev[2], PhD; Victor J Strecher[3], PhD, MPH; Kendra Rothert[4], MHS; Julia Anderson[5], MA

[1]Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

[2]RTI International, Research Triangle Park, NC, USA

[3]University of Michigan Center for Health Communications Research and HealthMedia, Inc, Ann Arbor, MI, USA

[4]Kaiser Permanente Care Management Institute, Oakland, CA, USA

[5]Group Health Cooperative, Seattle, WA, USA

Corresponding Author:
Mick P Couper, PhD
Institute for Social Research
University of Michigan
PO Box 1248
Ann Arbor, MI 48109
USA
Phone: +1 734 647 3577
Fax: +1 734 764 8263
Email: mcouper@umich.edu

Abstract

Background: Attrition, or dropout, is a problem faced by many online health interventions, potentially threatening the inferential value of online randomized controlled trials.

Objective: In the context of a randomized controlled trial of an online weight management intervention, where 85% of the baseline participants were lost to follow-up at the 12-month measurement, the objective was to examine the effect of nonresponse on key outcomes and explore ways to reduce attrition in follow-up surveys.

Methods: A sample of 700 nonrespondents to the 12-month online follow-up survey was randomly assigned to a mail or telephone nonresponse follow-up survey. We examined response rates in the two groups, costs of follow-up, reasons for nonresponse, and mode effects. We ran several logistic regression models, predicting response or nonresponse to the 12-month online survey as well as predicting response or nonresponse to the follow-up survey.

Results: We analyzed 210 follow-up respondents in the mail and 170 in the telephone group. Response rates of 59% and 55% were obtained for the telephone and mail nonresponse follow-up surveys, respectively. A total of 197 respondents (51.8%) gave reasons related to technical issues or email as a means of communication, with older people more likely to give technical reasons for noncompletion; 144 (37.9%) gave reasons related to the intervention or the survey itself. Mail follow-up was substantially cheaper: We estimate that the telephone survey cost about US $34 per sampled case, compared to US $15 for the mail survey. The telephone responses were subject to possible social desirability effects, with the telephone respondents reporting significantly greater weight loss than the mail respondents. The respondents to the nonresponse follow-up did not differ significantly from the 12-month online respondents on key outcome variables.

Conclusions: Mail is an effective way to reduce attrition to online surveys, while telephone follow-up might lead to overestimating the weight loss for both the treatment and control groups. Nonresponse bias does not appear to be a significant factor in the conclusions drawn from the randomized controlled trial.

(J Med Internet Res 2007;9(2):e16) doi:10.2196/jmir.9.2.e16

KEYWORDS

Nonresponse; attrition; Internet; weight management; randomized controlled trial

Introduction

Online interventions are an increasingly attractive method of reaching large numbers of potential participants for a wide variety of health interventions [1]. However, a key challenge of online health interventions is that of retaining subjects, especially for follow-up surveys to measure outcomes [2]. High rates of attrition, or dropout, can seriously threaten the inference from evaluations of online interventions, both in terms of external validity (generalizability) and internal validity [2, 3]. Online interventions appear to be particularly susceptible to problems of attrition. For example, only 1% of participants completed all 12 weeks of a panic disorder program [4], 33% completed all five modules of a depression program [5], and 35% completed a follow-up questionnaire about 10 weeks after enrollment in a smoking-cessation trial [6]. In his review of the problem, Eysenbach [2] argues for a "science of attrition," saying that "Nonusage data *per se* should be of great interest to researchers, and attrition curves may be underreported and underanalyzed."

While mixed-mode designs are increasingly common in survey research (for a review, see [7]), both as a way to reduce costs and to increase response rates, such mode switches are less common in health interventions, especially those involving online methods. In one exception, Tomson et al [8] reported on a telephone interview with nonresponders to a mail follow-up survey in a smoking intervention. Of the 84 subjects they followed up, 46 responded (for a 55% response rate). They found that those who did not respond to the original survey were more likely to be smoke-free at 12 months (39%) than the original mail respondents (31%). When asked about reasons for not returning the mail questionnaire, 35% claimed that they had returned it and 20%, that they had not received it.

While there are several comparisons of mail versus email for health surveys (eg, [9-12]), we are aware of no studies that have followed up nonrespondents to an online survey using an alternative mode of data collection. Crawford and colleagues [13] used telephone follow-up for nonrespondents to a mail and Web survey to remind people to complete the survey and to ascertain reasons for nonresponse, but not to collect follow-up data. Clarke and colleagues [14] assigned participants to a telephone or mail reminder, but similarly did not use these modes to collect follow-up data.

This paper represents an attempt to better understand the attrition problem in a randomized controlled trial (RCT) of an online weight management intervention and to find ways to counter the potential negative effect of attrition on the conclusions that can be drawn from such studies. We describe a follow-up procedure to examine why people drop out and what can be done about it. We explore alternative modes (mail and telephone) for following up nonresponders to the online surveys. Finally, we discuss the implications of this work for online interventions and follow-up surveys.

Methods

Background on the Online Intervention

This nonresponse follow-up study is part of a broader project aimed at evaluating tailored versus nontailored Web-based weight management materials. Details of the study have been described elsewhere [15]. Kaiser Permanente (KP) members in four regions of the United States were recruited using a variety of methods (through clinicians, member newsletters, and letters to members of diabetes and cardiovascular disease registries). A total of 4041 eligible participants were enrolled over a 6-month period beginning in September 2002. Eligible participants were current members of KP who were age 18 and older, had regular access to the Web and a functioning email address, had a body mass index (BMI) of 25 or greater, and who expressed a willingness to complete follow-up questionnaires. The average age of participants was 45.4 years (SD = 12.1); 82.8% were female, 56.6% were white, and 35.6% were African American. Participants had an average weight of 92.3 kg (SD = 14.4) and an average BMI of 32.1 (SD = 3.9).

Participants completed the baseline questionnaire online, following which they were randomized to one of two treatment arms: Web-based tailored ("expert system") weight management materials or Web-based nontailored ("user navigated") weight management materials, with the latter serving as the control group. Following the 6-week weight management program, participants were assessed by an Internet-based survey 3, 6, and 12 months after baseline assessment.

Email notices of the availability of each follow-up questionnaire were sent to participants, who received as many as 21 email reminders over a 3-week period before being considered as a nonrespondent to each follow-up assessment. All participants originally enrolled at baseline were sent email prompts to complete the Web-based 6-month and 12-month surveys, regardless of response status at earlier follow-up waves. No incentive was offered for completion of the online follow-up surveys.

Of the participants enrolled at baseline, 31% responded to the Web-based 3-month follow-up survey, while 21% responded to the 6-month survey, and 15% responded to the 12-month survey. There were no significant differences in attrition by baseline assignment to treatment arm, baseline BMI, or a variety of other baseline measures. However, given that 85% of the baseline participants were lost to follow-up at the 12-month measurement, it is important to explore the effect that this may have had on the results of this study. The goal of the present paper is to examine the reasons for loss to follow-up in the online weight management intervention and examine possible differences between those who were lost and those who were retained in the study.

Design and Implementation of the Nonresponse Follow-Up Survey

Given that only 21% of the original participants responded to the online 6-month survey, a small pilot study was conducted among nonrespondents to explore possible reasons for dropout. A telephone survey was used to contact 104 participants, of

whom 44 agreed to be interviewed and provide relevant information, with 42 providing weight information. The results of the 6-month follow-up are reported elsewhere [15]. No significant differences were found between 6-month respondents and the 6-month nonrespondent sample interviewed by telephone in terms of weight loss, weight, motivation to manage weight, self-efficacy, or program rating. Of the 44 interviewed, 20 (45%) reported not receiving any email notification for either the 3- or 6-month follow-up survey. The success of this pilot study led us to design a larger follow-up after the 12-month online survey.

One of the four KP regions participating in the weight management intervention was dropped because of administrative delays in approving the follow-up studies. The Institutional Review Boards at the remaining three KP regions, as well as Group Health Cooperative and the University of Michigan, approved the study protocol. This left us with a total of 3260 baseline participants—1681 treatment and 1579 control participants. The pattern of response to the three online follow-up surveys is shown in Table 1.

Table 1. Response pattern to three online follow-up surveys

Response Pattern			Treatment		Control		All	
3-month	6-month	12-month	No.	%	No.	%	No.	%
R	R	R	202	12.0	162	10.3	364	11.2
NR	R	R	21	1.2	21	1.3	42	1.3
R	NR	R	29	1.7	31	2.0	60	1.8
NR	NR	R	17	1.0	16	1.0	33	1.0
R	R	NR	115	6.8	113	7.2	228	7.0
NR	R	NR	34	2.0	29	1.8	63	1.9
R	NR	NR	177	10.5	192	12.2	369	11.3
NR	NR	NR	1086	64.6	1015	64.3	2101	64.4
Total			1681	100.0	1579	100.0	3260	100.0

R = respondent; NR = nonrespondent

As shown in Table 1, the majority of participants (64.4% overall) did not complete any of the follow-up surveys, while 499 (15.3%) completed the 12-month follow-up (whether or not they completed earlier follow-up surveys). The pattern of response across the three waves of follow-up is quite similar between the treatment and control groups.

The study reported here follows up on those who did not respond to the 12-month survey, regardless of their response to the 3- and 6-month surveys. This left us with 2761 participants eligible for the nonresponse follow-up study, 1412 from the treatment group and 1349 from the control group. Given that those who did not do any of the follow-up surveys comprised the largest group, we drew a systematic sample of cases from this group, but selected all those who did one or two of the follow-up surveys but not all three. In this way, we selected a total of 700 baseline participants, 350 from the treatment group and 350 from the control group. The sample size was determined largely by budget constraints.

We note that this is not an equal probability selection of nonrespondents. Those who did none of the follow-up surveys are under-represented in this sample relative to those who did one or more follow-ups. However, we used unweighted analyses because our focus was more on the differences between modes (see below) and differences between treatment arms than on the differences by pattern of nonresponse. However, we also conducted weighted analyses, and these led to similar conclusions as those presented here.

The rationale for the design of the 12-month nonresponse follow-up was based on several expectations:

- We hypothesized that much of the attrition may be due to reasons unrelated to a particular arm of the weight management intervention.
- We expected that a change in data collection mode may bring many of the nonrespondents back in.
- We wanted to evaluate the cost-effectiveness of alternative follow-up strategies.

In particular, while telephone follow-up is often an effective way of increasing response to other modes of data collection (especially mail), it is both more costly and raises concerns about social desirability and the effects of instrument design. The presence of an interviewer is known to affect reporting of socially sensitive information [16]. In addition, the visual versus aural presentation of survey questions may affect the answers obtained in the two modes [17]. For these reasons we were interested in the efficacy of a mail follow-up relative to a telephone follow-up. We expected the mail follow-up to be more similar to the Web in the measurement properties and to be cheaper than the telephone, but to take longer and be less effective than the telephone in gaining cooperation from nonrespondents to the online survey.

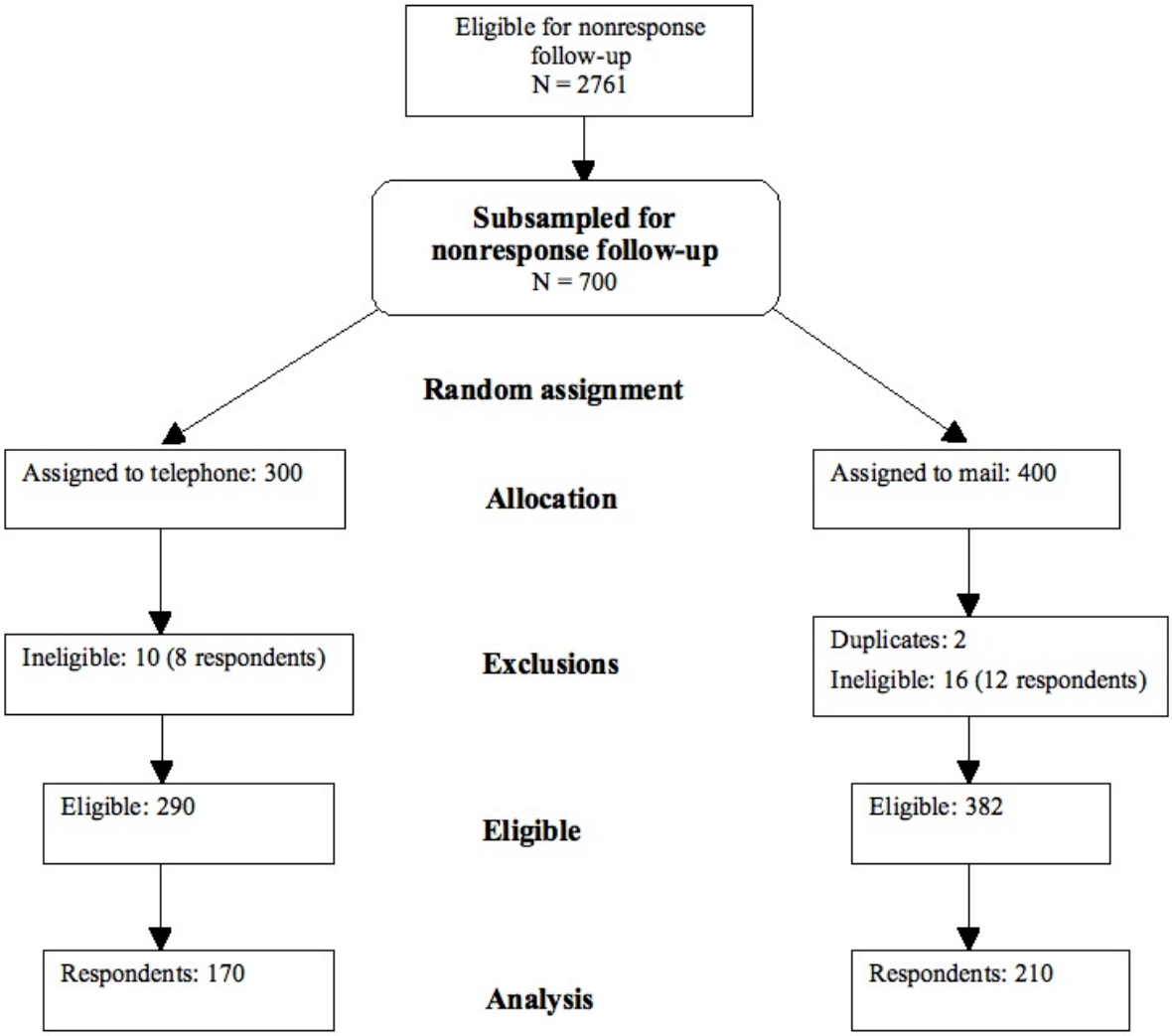
The use of the telephone for follow up is often predicated on the assumption that the online questionnaire was received, and that those who did not return it may need to be persuaded to participate. Switching from one self-administered mode (Web) to another (mail) is a largely untested approach. But if it works,

it has potential cost benefits as well as the advantage of greater comparability of measurement. For this reason, we embedded a mode experiment in the nonresponse follow-up study, with 300 of the nonrespondents being randomly assigned to telephone and 400, to mail. The disproportionate allocation reflects the cost differential between the two modes.

The mail follow-up survey involved a single mailing. Questionnaires were mailed with a KP return address and a cover letter signed by each of the three KP regional directors. Completed questionnaires were returned to Group Health Cooperative (GHC) for processing. Each mailed questionnaire included a US $5 bill as a token of appreciation for completing the questionnaire. The questionnaire was printed on a single 8-1/2" by 14" sheet folded in booklet form and contained a maximum of 13 questions to be answered. Two duplicates were discovered during the mailing process, leaving us with a sample of 398 mail cases.

The telephone survey was conducted by trained interviewers at GHC. No advance letter or incentive was used. Up to 15 call attempts were made on various days and at various times of day. The average interview length was 5.36 minutes. Figure 1 presents a flowchart of the nonresponse follow-up recruitment process, following the CONSORT model.

Figure 1. Nonresponse follow-up recruitment flowchart



Data from the mail and telephone survey were combined into a single data file and merged with the baseline and online follow-up responses for analysis. We first examine the results of the nonresponse follow-up survey before looking at correlates of attrition and nonresponse in both the 12-month online survey and the nonresponse follow-up survey.

Results

The Nonresponse Follow-Up Survey

Response Rates

Based on prior research, we anticipated a 40% response rate and expected the telephone mode to yield a higher response rate than mail. Overall, 400 out of 698 responded to the nonresponse follow-up, for a 57.3% response rate. Unfortunately, we learned

after the fact that 26 cases in our sample (including 20 respondents) were not eligible for the intervention, so the final count was 380 respondents. Subsequent analyses are based on these 380 respondents.

Of the 290 cases in the telephone sample, 170 responded, for a 58.6% response rate; for those in mail sample, 210 out of 382 responded, for a 55.0% response rate. These rates did not differ significantly (χ^2_1= 0.9, *P* = .34). However, the mail survey took substantially longer to complete, and, if time were a factor, the telephone mode would yield a higher response rate within 2 weeks, or even a month, after the start of follow-up. The median response time for the mail survey was 10 days (mean = 16.2, SD = 13.9), while that for the telephone survey was 8 days (mean = 8.9, SD = 6.5). If we had cut off data collection at 14 days, we would have obtained 61.9% of the mail respondents and 78.8% of the telephone respondents. Similarly, cutting off data collection at 28 days would acquire 85.7% of mail respondents and all but one of the telephone respondents. The last completed mail questionnaire was received some 6 months after the initial mailing.

Overall, the number of online follow-up surveys that participants completed is predictive of whether they completed the nonresponse follow-up survey: 50.1% of those who did not respond to any of the three online surveys completed the nonresponse survey, compared with 59.8% of those who did not respond to 2 of the 3 online surveys, and 65.3% of those who did not respond to only 1 of the 3 online surveys (χ^2_2= 9.0, *P* = .01). This pattern is similar for both the mail and telephone samples. But, even so, about half of those who did not complete any measurements following baseline were still brought back into the study some 12 months later with the nonresponse follow-up study.

Costs of Follow-Up

What are the relative costs of conducting mail versus telephone follow-up surveys of nonrespondents? To produce crude estimates, we took the total cost of each operation and divided by the sample size and the number of completes, respectively. Costs for the telephone included the development of the short computer-assisted telephone interviewing (CATI) instrument and conduct of the survey. Costs for the mail survey included the cost of printing the questionnaires and assembling the mailing packets, the cost of incentives and postage, and the cost of keying the data from the returned questionnaires.

Based on these rough numbers, we estimate that the telephone survey cost about US $34 per sampled case, compared to US $15 for the mail survey. On the basis of completed questionnaires, the relative costs were approximately US $57 per completed case for the telephone survey and US $28 for the mail survey. Given this cost differential, the mail survey was not only almost as effective in terms of response rate as the telephone survey, but also substantially less expensive.

Reasons for Nonresponse

We included a few questions in the nonresponse follow-up survey to ascertain reasons for earlier nonresponse. Based on the results of the 6-month pilot follow-up, these questions primarily focused on technical barriers to completing the online surveys.

We first asked whether participants recalled receiving the email message for the 12-month online follow-up survey. If they had, we asked if they recalled reading the message. If they did so, we went on to ask if they clicked on the link to access the survey. For those who did not recall receiving an email message, we asked them to provide their current email address. The results from this series of questions are presented in Table 2.

Table 2. Responses to nonresponse follow-up questions about the email invitations

	Mail (n=210)	Telephone (n=170)	Total (n=380)
Recall receiving email message	53.8% (113/210)	43.5% (74/170)	49.2% (187/380)
Recall reading email message	44.3% (93/210)	28.8% (49/170)	37.4% (142/380)
Among those not recalling receiving the email message:			
Do you still have access to email? (% yes)	87.6% (85/97)	96.9% (93/96)	92.2% (178/193)
Among those not recalling reading the email message:			
How often do you typically check email? (% at least 1-2 times per week)	73.5% (86/117)	71.9% (87/121)	72.7% (173/238)
How often do you delete email without opening? (% at least sometimes)	65.0% (76/117)	65.3% (79/121)	65.1% (155/238)
Among those recalling receiving the email message:			
Did you click on the link to access the survey? (% yes)	47.8% (54/113)	31.1% (23/74)	41.2% (77/187)

Mail respondents reported higher rates of recall for both receiving and reading the email message than did telephone respondents. One reason may be that the mail mode gives respondents more time to consider the question and retrieve information from memory, or even check their email inbox.

An additional 13.8% (36.4% of mail respondents and 16.7% of phone respondents) who answered “no” to the question on recalling receiving the email message still answered that they clicked on the link to the survey. Thus, 26.1% of respondents overall (29.5% of mail and 21.8% of phone) reported clicking on the link for the survey, regardless of whether they recalled receiving or reading the email message.

All respondents to the nonresponse follow-up who said they still had access to email, whether or not they recalled receiving or reading the email invitations, were asked in an open-ended question the main reason for not accessing the website or completing the survey. The responses were classified into several themes, summarized in Table 3. The responses did not appear to differ by mode, so the combined distribution of responses is presented.

Table 3. Classification of reasons for noncompletion of the online survey into major themes

Theme	No.	%
No response to question	21	5.5
Technical problems		
Problems accessing or submitting the survey	59	15.5
Email or computer problems (including those with no access to email)	35	9.2
Did not receive or remember message, or treated it as SPAM	103	27.1
Problems with intervention or survey		
Lack of interest in or lack of effectiveness of intervention	67	17.6
Survey was boring or too long	8	2.1
Medical or personal reasons	8	2.1
Refused to do survey	8	2.1
No time or bad timing	53	13.9
Other		
Don’t remember	5	1.3
Did complete the survey	13	3.4
Total	380	100.0

By far the largest number of nonresponse follow-up respondents (103) gave reasons related to not recalling receipt of the email message. If we treat this as a technical problem, we see that a total of 197 respondents (51.8%) gave reasons related to technical issues or email as a means of communication. In contrast, 144 (37.9%) gave reasons related to the intervention or the survey itself (see Table 3). This suggests that the medium of communication (email) may be as much if not more of a factor in attrition as the intervention. This is one of the reasons we believe that the mail and telephone follow-up surveys were successful. Of course, these are post hoc justifications of why participants did not do the online survey, and we don’t know the reasons for noncompletion for those participants who did not respond to the nonresponse follow-up survey.

To explore whether some types of participants may be more likely to report technical problems than others, we collapsed the reasons in Table 3 into three broad groups (technical, substantive, and other) and ran a multinomial logistic regression using baseline demographic and design variables as predictors. Only age was significantly (P = .03) associated with the type of reason given, with older people more likely to give technical reasons for noncompletion of the 12-month online survey: 43% of those 45 years and younger cited technical problems, compared to 55% of those 46-60 and 67% of those 61 or older.

Differential Response by Mode

While we found no significant differences in the overall response rate to the nonresponse follow-up by mode, are some people more likely to respond to one follow-up mode than the other? While the response rates are similar, it could be that the mail follow-up brings in different types of people than the telephone follow-up survey. To examine this, we ran a logistic regression model predicting whether a case was a mail or telephone respondent, conditional on having completed the follow-up survey. None of the demographic or design variables we examined—age, gender, race, education, KP region, or treatment group—was statistically significant. This suggests that the differences in the answers to the survey questions by mode (see below) are likely due to the features of the modes themselves rather than differential selection into the two groups.

Mode Effects

Several rating scales were included in the follow-up questionnaires. Based on the literature (eg, [18]), we expected two types of response effects. First, response options presented early in a list are more likely to be selected in visual presentation modes (mail and Web), while those presented later in a list are more likely to be selected in aural modes (telephone). In others words, primacy effects are likely in the mail mode, while recency effects are more likely in the telephone mode [19]. The second effect is that of socially desirable responding, with more positive (ie, socially desirable) responses expected on the telephone given the presence of an interviewer.

Four of the ratings were presented with the most negative response (eg, not at all confident, not at all satisfied, not at all motivated) listed first. The responses were read in the same order on the telephone. The response order effect (primacy in mail, recency in telephone) is expected to produce lower means (more negative) for the mail respondents, as would the social desirability effect. In other words, the two effects are expected to reinforce each other. The means and standard deviations for these items are presented in Table 4. For three of the four items, we find significantly lower means for mail than telephone: confidence in managing one's own weight (t_{341} = 2.64, P = .009), confidence in maintaining recommended levels of physical activity (t_{352} = 3.54, P < .001), and motivation to manage one's own weight (t_{371} = 2.66, P = .008). For the fourth item of this type, a rating of satisfaction with care at KP, no differences were found (t_{366} = 0.34, P = .73).

One item, a rating of the online weight management information, was ordered from the most positive (excellent) to the least positive response (poor). Here, the response order and social desirability effects are expected to cancel each other out. First, nonresponse to this item was higher than all other items (20% of mail respondents and 25% of phone respondents did not answer), perhaps reflecting the fact that some participants may not have spent much time with the online materials. This is borne out when we look at the treatment group only: of those who did not look at any of the online newsletters, 27.7% did not answer this question, compared to 8.1% for those who looked at one or more newsletters. Among those who did answer the question, we find no significant differences in the mean ratings on this item by mode (t_{264} = 0.33, P = .74).

Table 4. Mean responses (SD) to nonresponse follow-up measures, by mode

Question	Mail	Telephone
Q4. Overall, how would you rate the online weight management information that you received? (1 = Excellent to 5 = Poor)	2.97 (1.02)	2.93 (1.08)
Q5. Currently, how confident are you that you can manage your weight? (1 = Not at all to 5 = Extremely)	2.05 (0.97)	2.34* (1.09)
Q6. Using a scale from 0 to 10, where 0 means not at all motivated and 10 means extremely motivated, how motivated are you to manage your weight?	6.05 (2.57)	6.72* (2.35)
Q7. How confident are you that you can maintain recommended levels of physical activity? (1 = Not at all to 5 = Extremely)	2.36 (1.09)	2.76* (1.14)
Q8. How much do you currently weigh? (pounds)	209.6 (46.61)	208.6 (46.06)
Q9. Finally, how satisfied are you overall with your care at Kaiser Permanente? (1 = Not at all to 5 = Extremely)	3.60 (1.00)	3.56 (0.94)

*P < .05, comparing mail versus telephone (t-test, see text for P-values)

Nonresponse to the most critical question ("How much do you currently weigh?") did not differ by mode, with 8.1% of mail and 8.8% of telephone respondents not providing an answer to this question. Among those who did answer, the telephone respondents reported a lower weight on average than the mail respondents; however, this did not reach statistical significance (t_{332} = 0.19, P = .85; see Table 4). If we examine reported weight loss from baseline to the 12-month nonresponse follow-up survey (the key dependent variable), we find significant effects, with the telephone respondents reporting greater weight loss than the mail respondents, whether reported in kilograms (an average weight loss of 3.30 kg for telephone respondents and 1.19 kg for mail respondents; t_{320} = 3.1, P = .002) or BMI (an average BMI reduction of 1.20 for telephone respondents and 0.45 for mail respondents; t_{315}= 2.96, P = .003). Given the known social desirability effects associated with the telephone, we believe that the mail responses are more "honest" than those provided over the telephone. This is consistent with the view in the mode effects literature (eg, [17,18]) that higher reports of socially undesirable behaviors or attributes (and lower reports of socially desirable ones) reflect greater accuracy of reporting. Our findings suggest that if we had done a telephone follow-up only, we would have overestimated the weight loss for both the treatment and control groups (the mode difference does not interact with experimental condition). Given that the mail mode is more similar to the online measurement used for both baseline and follow-up surveys, we believe that the smaller weight loss estimated for the mail respondents more closely reflects the truth.

The evidence for social desirability bias in the telephone responses echoes findings from other studies. For example, Eicheldinger et al [20] conducted a follow-up of nonrespondents

to the Consumer Assessment of Health Plans Study (CAHPS), randomly assigning participants to telephone or mail (using overnight delivery) follow-up. While their response rates were lower than ours (23.7% for mail and 27.1% for phone), they found that those who responded by telephone were more likely to report the most positive response to 13 of the 20 performance measures. Similarly, in a study of employees at a large company who were randomly assigned to Web or telephone modes of data collection [21], significant differences were obtained for mean satisfaction with the health insurance plan (6.88 for Web and 7.32 for telephone, $P < .05$) and for mean self-rated health (3.51 for Web and 3.79 for telephone, $P < .01$). Similar effects are found in comparisons of mail versus telephone [22] and Web versus telephone [23]. However, our results suggest that it may not just be social desirability effects at work; differences in format or layout of the items may also produce mode effects [24].

Modeling Nonresponse

In addition to data from the nonresponse follow-up survey, we also have information on all participants from the baseline survey. In this section, we use these data to examine correlates of nonresponse to the 12-month online survey and also to the mail and telephone nonresponse follow-up surveys.

In contrast to cross-sectional sample surveys in which little is known about sample members who do not participate, one of the advantages of an (online) intervention such as this is that a lot of information may be collected at baseline, and these data can be used to examine who drops out and who does not, among those who enrolled. The data can be used not only to examine patterns of differential attrition among subgroups, but also to statistically adjust for such patterns at the time of analysis.

We ran several logistic regression models, first predicting response or nonresponse to the 12-month online survey, then, predicting response or nonresponse to the follow-up survey (among those included in the eligible for nonresponse follow-up sample). We briefly summarize these models below.

Response to the 12-Month Online Survey

The first model included the following baseline demographic and design variables: age, gender, race, education, KP region, BMI, study assignment (treatment or control). The coefficient of determination, R^2, [25] for this model was 0.030, while the Nagelkerke [26] max-rescaled R^2 measure was 0.052 (see [27] for a discussion of alternative pseudo R^2 indices). This suggests that these baseline variables do not do a very good job of predicting whether a participant will be a respondent or nonrespondent to the 12-month online survey. This is reassuring in that the attrition does not appear to vary much by these baseline characteristics. Specifically, the dropout rates for the treatment and control groups did not differ significantly. There were no significant differences in 12-month completion by baseline BMI or by gender. The KP regions differed significantly in their 12-month completion rates, ranging from 10.2% to 18.8% across the three regions in the study, but there was no significant interaction with treatment group. Age was associated with a significant ($P = .01$) positive effect on completion. Minorities (African Americans and those of other races) were significantly ($P = .008$) less likely to complete the 12-month survey, as were those with lower levels of education ($P = .009$). Both of these variables are correlated with lower levels of Internet access and may be associated with greater risk of losing such access over the life of the study [28,29].

To this model, we added a set of behavioral and attitudinal measures related to the online intervention from the baseline survey. These included whether the participant had received medical advice to lose weight, how successful they were at losing weight in the past, their weight loss goals for the program, their motivation for losing weight, frequency of exercise and physical activity, self-rated health, and satisfaction with KP. The addition of these variables did not significantly improve the fit of the model ($\chi^2_{14} = 30.7$ [for Δ in -2 LogL]), producing a max-rescaled R^2 of 0.068. Among the added variables, only level of physical activity ($P = .03$), with those doing light exercise being more likely to complete the survey than those doing moderate to heavy activity, and self-rated health ($P = .009$), with better health associated with higher rates of completion, were statistically significant under the model. In other words, baseline measures of motivation to lose weight, weight loss target, satisfaction with KP, assignment to treatment or control group, and the like, were not significantly associated with completion of the 12-month online survey. This provides some reassurance that nonresponse bias may not be large—at least in terms of variables measured at baseline.

A final model added a set of process measures from the intervention, namely whether the participant completed the 3- and 6-month online surveys. As expected, nonresponse to one of the early follow-up surveys was highly predictive of nonresponse to the 12-month follow-up survey, with conditional odds ratios of 4.0 (95% CI, 2.90-5.53) and 13.1 (95% CI, 9.78-17.62) for completion of the 3- and 6-month survey, respectively. In addition, those in the treatment group were given access to three online newsletters as part of the intervention. Among this group, the number of newsletters they accessed on the website was predictive of 12-month survey completion. The odds ratio of being a respondent at 12 months for those who opened no newsletters was 0.14 (95% CI, 0.07-0.26) relative to those who accessed all three, while for those who accessed one newsletter, it was 0.49 (95% CI, 0.29-0.83), and for those who accessed two newsletters, it was 0.52 (95% CI, 0.32-0.86). These limited process indicators suggest two conclusions: (1) those who are actively engaged in the intervention (ie, who show evidence of visiting the website and accessing material) have higher completion rates, and (2) those who responded to earlier follow-up surveys are more likely to respond to the final (12-month) follow-up survey.

These conclusions, in turn, have two implications. First, online interventions can provide researchers with a wide variety of measures of active engagement in the program [30]. These indicators can include number of sessions logged in, time spent online, number of pages viewed, and so on. Such process data or paradata [31] can be routinely captured as part of such online interventions and can be useful not only for understanding how much time and attention is spent on different parts of the website (with a view to identifying areas for improvement), but also as

a measure of how much participants are being exposed to the stimulus material. This could serve as an important mediator variable in analyses of various outcome measures. Second, when multiple follow-ups are part of the design, nonresponse to earlier follow-up surveys can identify participants at risk for dropout, permitting researchers to target intervention strategies aimed at retaining such participants in the study. The responsive design strategies being developed to reduce nonresponse in surveys (eg, [32]) can similarly be deployed to counter nonresponse in online interventions. Online studies not only permit targeted or tailored interventions, but also tailoring of data collection and follow-up strategies.

Response to the Nonresponse Follow-Up Survey

The second set of models parallels the first, but focuses on completion of the mail or telephone follow-up survey, among those included in the nonresponse follow-up study. These models are based on 672 eligible participants included in the nonresponse follow-up, 380 of whom completed either the mail or telephone survey. The max-rescaled R^2 measure for the demographic and design variables was 0.067. Only age remained a significant ($P < .001$) predictor of response to the nonresponse follow-up survey, with older people more likely to complete the survey. Interestingly, while African Americans were less likely than whites to complete the 12-month online survey, they appeared slightly but not significantly ($P = .36$) more likely (OR = 1.25, 95% CI, 0.85-1.85) to complete the mail or telephone follow-up survey. Similar effects are found for those with high school or lower education relative to those with a college degree ($P = .59$, OR = 1.3, 95% CI, 0.75-2.26). This may provide some support for the observation that those at greatest risk of losing access to the Internet (older persons, minorities, those with lower education) may be brought back into the analytic sample with alternative modes of data collection.

As before, adding the baseline behavioral and attitudinal measures to this model did not improve the fit, relative to the base model. Only one of the added predictors was statistically significant ($P < .05$), with those who spend 2 hours or more a day in front of the TV or computer outside of work being less likely to respond to the follow-up survey ($P = .02$). This suggests that the decision to complete the nonresponse follow-up survey is made largely independent of the original decisions regarding participation in the online intervention.

Adjusting for Nonresponse

One of the goals of conducting the nonresponse follow-up survey among a sample of nonrespondents, in addition to exploring reasons for nonresponse, was to obtain data to inform statistical adjustment for nonresponse. In his review of the attrition problem, Eysenbach [2] argues that an intent-to-treat analysis, in which all dropouts are assumed to have negative or neutral outcomes, is the only way to avoid selection bias. We argue rather that weighting or imputation based on informed models of attrition or dropout requires fewer assumptions about the missing cases. As Hollis and Campbell [33, p. 673-674] note, "…no imputation method can give an unbiased estimate of the treatment effect unless the assumptions made about the missing data are valid." A nonresponse follow-up study allows one not only to reduce the amount of missing data, but also to evaluate the missing data assumptions. To quote Hollis and Campbell [33, p. 674] again, "To fully appreciate the potential influence of missing responses, some form of sensitivity analysis is recommended, examining the effect of different strategies on the conclusions."

In work described elsewhere [34], we used the data from the nonresponse follow-up to multiply-impute data for the remaining 12-month online nonrespondents [35]. This method utilizes all available data, while accounting for the uncertainty due to imputation. Using a complete case analysis from the 12-month online respondents only, we would reach a conclusion that the treatment had a statistically significant effect on weight loss at 12 months relative to the control. However, using the data from the nonresponse follow-up to impute the missing 12-month cases, we would conclude that the differences between treatment and control, although still in the expected direction, do not reach statistical significance. These models are limited by the small number of cases included in the nonresponse follow-up relative to the number of nonrespondent cases and by the differences we found between the two modes of follow-up. Therefore, these results can only be viewed as suggestive. However, they allow us to explore the sensitivity of the substantive models to different assumptions about the missing data at the 12-month follow-up.

Discussion

Our study has several potential limitations. First, the nonresponse follow-up was conducted within the context of a weight management intervention, which was restricted to overweight and obese members of a health maintenance organization (HMO) with regular Internet access. This may limit generalizability to other populations and settings. Second, this study did not test different ways to enhance the response rate to the 12-month online survey (eg, by using incentives). The success of the follow-up effort may be conditional on the initial response obtained. Third, this was a small-scale exploratory study embedded in a larger study. The small sample size may limit our ability to draw statistically reliable conclusions.

Nonetheless, we have learned a number of things from this exploratory study. First, a significant proportion of those who drop out of an online RCT or intervention can be brought back by switching modes of data collection. A variety of technical reasons, unrelated to the online intervention, can account for a substantial proportion of such dropout, and modes switches are an effective counter to the uncertainties of using email as a communication medium.

Second, we learned that mail is almost as effective as the telephone for such follow-up. Further, it is significantly cheaper, and it is more similar to the original online mode in terms of visual design and response styles and shares the absence of social desirability effects associated with interviewers. For these reasons, we believe that the mail survey produced responses that are more comparable to the online responses than did the telephone survey. Telephone calls can be a useful tool for prompting or reminding respondents to return their

questionnaires, but we believe that mail is a cost-effective method of following up online nonrespondents, if time is not a critical factor. On the basis of this work, we implemented a mail-only nonresponse follow-up study in a second controlled trial of a weight loss program [30].

Third, nonresponse follow-up studies such as this not only increase the number of cases for analysis but also help us understand the differences between those who drop out and those who complete all follow-up surveys. In other words, our analyses of treatment effects are not forced to rely on the often-heroic assumptions required by complete-case analysis. Nonresponse, or attrition, bias can be reduced in two ways: one is to reduce the rate of attrition, and the other is to reduce or measure the differences between those who drop out and those who don't [36]. We believe that following up nonrespondents—whether a sample of them as we did here, or all nonrespondents—using a different mode is a cost-effective way of increasing the analytic power and reducing the potential bias that may result from the relatively high rates of dropout experienced in online interventions and follow-up surveys.

Acknowledgments

Support for the evaluation of the tailored weight management study (THeME) came from Kaiser Permanente, and the study was implemented by Health Media, Inc. Support for the nonresponse follow-up study came from a development grant from the University of Michigan's Center for Health Communications Research (PI Victor J Strecher), with primary funding from the National Cancer Institute. Data collection for the nonresponse follow-up was conducted by Group Health Cooperative. Staff at HealthMedia, Inc, particularly Al Zielke, provided help with selection of the sample, provision of data, and with the follow-up study.

Conflicts of Interest

Dr. Strecher is Founder, Chairman, and Chief Science Officer of HealthMedia, Inc, which developed and has proprietary interest in the tailored weight management program described herein.

References

1. Evers KE, Cummins CO, Prochaska JO, Prochaska JM. Online health behavior and disease management programs: are we ready for them? Are they ready for us? J Med Internet Res 2005 Jul 1;7(3):e27 [FREE Full text] [Medline: 15998618] [doi: 10.2196/jmir.7.3.e27]
2. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [Medline: 15829473] [doi: 10.2196/jmir.7.1.e11]
3. Cook TD, Campbell DT. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston, MA: Houghton Mifflin; 1979.
4. 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]
5. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46 [FREE Full text] [Medline: 15631970] [doi: 10.2196/jmir.6.4.e46]
6. 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]
7. De Leeuw ED. To mix or not to mix data collection modes in surveys. J Off Stat 2005;21(2):233-255.
8. Tomson T, Björnström C, Gilljam H, Helgason A. Are non-responders in a quitline evaluation more likely to be smokers? BMC Public Health 2005 May 23;5(1):52 [FREE Full text] [Medline: 15910682] [doi: 10.1186/1471-2458-5-52]
9. Bälter KA, Bälter O, Fondell E, Lagerros YT. Web-based and mailed questionnaires: a comparison of response rates and compliance. Epidemiology 2005 Jul;16(4):577-579. [Medline: 15951679] [doi: 10.1097/01.ede.0000164553.16591.4b]
10. Harewood GC, Yacavone RF, Locke GR, Wiersema MJ. Prospective comparison of endoscopy patient satisfaction surveys: e-mail versus standard mail versus telephone. Am J Gastroenterol 2001 Dec;96(12):3312-3317. [Medline: 11774942] [doi: 10.1111/j.1572-0241.2001.05331.x]
11. Jones R, Pitt N. Health surveys in the workplace: comparison of post, email and World Wide Web methods. Occup Med 1999;49(8):556-558. [doi: 10.1093/occmed/49.8.556]
12. Leece P, Bhandari M, Sprague S, Swiontkowski MF, Schemitsch EH, Tornetta P, et al. Internet versus mailed questionnaires: a randomized comparison (2). J Med Internet Res 2004 Sep 24;6(3):e30 [FREE Full text] [Medline: 15471756] [doi: 10.2196/jmir.6.3.e30]
13. Crawford SD, McCabe SE, Couper MP, Boyd CJ. From Mail to Web: Improving Response Rates and Data Collection Efficiencies. Paper presented at: International Conference on Improving Surveys; August 2002; Copenhagen. .

14. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, Debar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005;7(2):e16 [FREE Full text] [Medline: 15998607] [doi: 10.2196/jmir.7.2.e16]
15. 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]
16. Groves RM, Fowler FJJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R. Survey Methodology. New York: Wiley; 2004.
17. Fricker S, Galesic M, Tourangeau R, Yan T. An experimental comparison of Web and telephone surveys. Public Opin Q 2005;69(3):370-392. [doi: 10.1093/poq/nfi027]
18. Tourangeau R, Rips L, Rasinski K. The Psychology of Survey Response. Cambridge, UK: Cambridge University Press; 2001.
19. Schwarz N, Hippler HJ, Noelle-Neumann E. A cognitive model of response order effects in survey measurement. In: Schwarz N, Sudman S, editors. Context Effects in Social and Psychological Research. New York: Springer-Verlag; 1992:187-201.
20. Eicheldinger CR, Iannacchione VG, Carpenter LC. Rating Medicare: does mode matter? In: Proceedings of the American Statistical Association, Survey Research Methods Section. Alexandria, VA: American Statistical Association; 2003:1318-1324.
21. Speizer H, Greene J, Baker R, Wiitala W. Telephone and Web: the Mixed-Mode Challenge. Paper presented at: Second International Conference on Telephone Survey Methods; Miami, FL. 2006.
22. Beebe TJ, Mcrae JA, Harrison PA, Davern ME, Quinlan KB. Mail surveys resulted in more reports of substance use than telephone surveys. J Clin Epidemiol 2005 Apr;58(4):421-424. [Medline: 15868697] [doi: 10.1016/j.jclinepi.2004.10.007]
23. Roster CA, Rogers RD, Albaum G, Klein D. A comparison of response characteristics from Web and telephone surveys. Int J Market Res 2004;46(3):359-373.
24. Dillman DA, Christian LM. Survey mode as a source of instability in responses across surveys. Field Methods 2005;17(1):30-52. [doi: 10.1177/1525822X04269550]
25. Cox DR, Snell EJ. The Analysis of Binary Data. 2nd edition. London, UK: Chapman and Hall; 1989.
26. ; Nagelkerke NJD. A note on a general definition of the coefficient of determination. Biometrika 1991;78(3):691-692. [doi: 10.1093/biomet/78.3.691]
27. Lemeshow S. Applied Logistic Regression. 2nd edition. New York: Wiley; 2000.
28. Fox S. Digital Divisions. Washington, DC: Pew Internet & American Life Project; 2005. URL: http://www.pewinternet.org/pdfs/PIP_Digital_Divisions_Oct_5_2005.pdf [WebCite Cache ID 5PZALUPHG]
29. ; National Telecommunications and Information Administration (NTIA). Falling Through the Net: Toward Digital Inclusion. Washington, DC: US Department of Commerce; 2000. URL: http://search.ntia.doc.gov/pdf/fttn00.pdf [WebCite Cache ID 5PZAsUGv0]
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]
31. Couper MP, Lyberg LE. The Use of Paradata in Survey Research. Proceedings of the 55th Session of the International Statistical Institute; Sydney, Australia. 2005.
32. Groves RM, Heeringa SG. Responsive design for household surveys: tools for actively controlling survey nonresponse and costs. J R Stat Soc Ser A 2006;169(3):439-457. [doi: 10.1111/j.1467-985X.2006.00423.x]
33. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 1999 Sep 11;319(7211):670-674 [FREE Full text] [Medline: 10480822]
34. Couper MP, Peytchev A, Little RJA, Rothert K. Combining Information from Multiple Modes to Evaluate and Reduce Nonresponse Bias. Paper presented at: Joint Statistical Meetings of the American Statistical Association; Minneapolis, MN. 2005.
35. Little RJA, Rubin DA. Statistical Analysis with Missing Data. 2nd edition. New York: Wiley; 2002.
36. Groves RM, Couper MP. Nonresponse in Household Interview Surveys. New York: Wiley; 1998.

Abbreviations

BMI: body mass index
GHC: Group Health Cooperative
KP: Kaiser Permanente
RCT: randomized controlled trial

submitted 06.11.06; peer-reviewed by T Beebe; comments to author 21.12.06; revised version received 03.02.07; accepted 08.05.07; published 13.06.07.

Please cite as:
Couper MP, Peytchev A, Strecher VJ, Rothert K, Anderson J
Following Up Nonrespondents to an Online Weight Management Intervention: Randomized Trial Comparing Mail versus Telephone
J Med Internet Res 2007;9(2):e16
URL: http://www.jmir.org/2007/2/e16/
doi:10.2196/jmir.9.2.e16
PMID:17567564

Original Paper

Reach, Engagement, and Retention in an Internet-Based Weight Loss Program in a Multi-Site Randomized Controlled Trial

Russell E Glasgow[1], PhD; Candace C Nelson[1], MA; Kathleen A Kearney[2], PhD; Robert Reid[3], MD, PhD; Debra P Ritzwoller[1], PhD; Victor J Strecher[4], PhD, MPH; Mick P Couper[4], PhD; Beverly Green[3], MD, MPH; Kevin Wildenhaus[5], PhD

[1]Kaiser Permanente Colorado, Denver, CO, USA
[2]Kaiser Permanente Care Management Institute, Oakland, CA, USA
[3]Group Health Cooperative, Seattle, WA, USA
[4]University of Michigan, Ann Arbor, MI, USA
[5]HealthMedia, Inc, Ann Arbor, MI, USA

Corresponding Author:
Russell E Glasgow, PhD
Kaiser Permanente Colorado
335 Road Runner Road
Penrose, CO 81240
USA
Phone: +1 719 372 3165
Fax: +1 719 372 6395
Email: russg@ris.net

Abstract

Background: Research increasingly supports the conclusion that well-designed programs delivered over the Internet can produce significant weight loss compared to randomized controlled conditions. Much less is known about four important issues addressed in this study: (1) which recruitment methods produce higher eHealth participation rates, (2) which patient characteristics are related to enrollment, (3) which characteristics are related to level of user engagement in the program, and (4) which characteristics are related to continued participation in project assessments.

Methods: We recruited overweight members of three health maintenance organizations (HMOs) to participate in an entirely Internet-mediated weight loss program developed by HealthMedia, Inc. Two different recruitment methods were used: personal letters from prevention directors in each HMO, and general notices in member newsletters. The personal letters were sent to members diagnosed with diabetes or heart disease and, in one HMO, to a general membership sample in a particular geographic location. Data were collected in the context of a 2×2 randomized controlled trial, with participants assigned to receive or not receive a goal setting intervention and a nutrition education intervention in addition to the basic program.

Results: A total of 2311 members enrolled. Bivariate analyses on aggregate data revealed that personalized mailings produced higher enrollment rates than member newsletters and that members with diabetes or heart disease were more likely to enroll than those without these diagnoses. In addition, males, those over age 60, smokers, and those estimated to have higher medical expenses were less likely to enroll (all $P < .001$). Males and those in the combined intervention were less likely to engage initially, or to continue to be engaged with their Web program, than other participants. In terms of retention, multiple logistic regressions revealed that enrollees under age 60 ($P < .001$) and those with higher baseline self-efficacy were less likely to participate in the 12-month follow-up ($P = .03$), but with these exceptions, those participating were very similar to those not participating in the follow-up.

Conclusions: A single personalized mailing increases enrollment in Internet-based weight loss. eHealth programs offer great potential for recruiting large numbers of participants, but they may not reach those at highest risk. Patient characteristics related to each of these important factors may be different, and more comprehensive analyses of determinants of enrollment, engagement, and retention in eHealth programs are needed.

(J Med Internet Res 2007;9(2):e11) doi:10.2196/jmir.9.2.e11

KEYWORDS

Internet; weight loss; recruitment; representativeness; retention; attrition; adherence; behavior change; randomized controlled trial; consumer health informatics

Introduction

There have been recent encouraging reports about the efficacy of Internet-based weight loss interventions [1-3], but many questions remain about the appeal and applicability of eHealth programs in real-world settings [4-6]. Specifically, it appears that computer-assisted and Internet programs may be effective for reducing fat intake [7] and increasing weight loss when supplemented by electronic health coaches [2,8,9].

One of the issues in need of greater understanding is recruitment to and participation in Internet-based health promotion programs [10]. We need to understand both the characteristics and motivation of those who participate versus those who decline, as well as the effects of different types of recruitment strategies. There has been considerable work on recruitment methods for clinical trials with in-person visits, especially on recruiting minority participants [11,12], but there is little controlled research on recruitment to Internet-based programs [13]. The ability to efficiently recruit and serve large numbers of consumers in remote locations is one of the key potential advantages of Internet-based programs, but better documentation and detailed understanding of this issue are needed.

An emerging issue of concern for Internet-based health behavior change interventions is high levels of attrition [1,14-16]. Participants use Internet-based resources differently than they do other modalities such as group and in-person programs, and for Internet programs serving large numbers of participants that do not heavily screen users or conduct studies under efficacy conditions, it has proven difficult to obtain follow-up information on a high percentage of initial participants [1,14-16]. At present, it is unclear what patient characteristics are related to active engagement with Internet program content and to participation in follow-up assessments and whether these characteristics are similar or different.

Based on previous promising work, a tailored online behavioral weight loss program (HealthMedia's Balance) was evaluated in a randomized controlled trial with and without the addition of extended goal setting and feedback and more intensive nutritional information. A prior multi-site randomized study of 2862 participants found the basic Balance program to produce significantly greater weight loss than the control condition [1].

The purposes of this paper are to (1) report on enrollment rates for an innovative, large-scale, Internet-based weight loss study, (2) analyze levels of program engagement and retention at 12-month follow-up, and (3) investigate recruitment method, setting, and patient characteristics associated with enrollment, program engagement, and retention in follow-up assessment.

Methods

Settings and Recruitment

The data reported here are part of a larger study assessing the impact of adding a nutrition component, a goal-setting component, or both to a tailored online weight management program (HealthMedia's Balance). The study was conducted in three prepaid group practice HMOs—the Ohio and Colorado regions of Kaiser Permanente, and Group Health Cooperative with enrollees from Washington state and Idaho. Health plan members with and without chronic illnesses (diabetes and coronary artery disease) were invited to participate, either through personal letters from medical leaders or through notices in general member communications such as HMO newsletters, flyers, and posters. Recruitment began in March 2004 and continued through early December 2004. Approval was obtained from Institutional Review Boards in each health plan. Interested members were directed to a study website where they completed a baseline assessment, reported their height and weight, learned whether they met the study criteria, and gave informed consent. Potential participants were excluded from the study if they had a body mass index (BMI) below the study minimum (< 30 for general membership and < 25 for those with chronic illness), were using drugs or surgery for weight loss, were participating in another organized weight loss program, or were physically or medically unable to exercise. Members excluded for these reasons were offered the opportunity to receive the basic Balance program if they wished. Respondents were also told that the Balance program was not intended for pregnant women, for those with eating disorders, or for those who had been diagnosed with heart failure; members who indicated that any of these categories applied to them were excluded.

Participants who met the enrollment criteria and consented to the study were randomly assigned within HMO to one of four interventions composed of combinations of three HealthMedia programs: (1) the 6-week Balance weight loss program alone; (2) Balance plus the 8-week nutrition management module Nourish; (3) Balance plus a simultaneous goal-setting component called Achieve; or (4) Balance plus Nourish and Achieve. Emails asking participants to complete follow-up surveys were sent to all participants 3, 12, and 18 months after enrollment. Participants were informed at the outset of the study that they would receive a US $10 gift certificate from Amazon.com or a similar online vendor each time they completed a follow-up questionnaire. Although not the purpose of this paper, for context we note that participants in all interventions were successful at losing weight.

Because of the low participation rate in follow-up assessments, a sample of nonrespondents to the 12-month online survey who provided mailing addresses were sent a printed survey by mail, along with US $10 cash. Of the 1796 nonrespondents, 913 were sent a mail survey; of these, 586 returned the completed survey, for a 64% response rate to the mail follow-up. (This compares to 56% for a mailed follow-up and 59% for a telephone follow-up in the earlier evaluation of the Balance program [15].) This follow-up effort boosted the unweighted 12-month response rate from 22.3% to 47.6% and the weighted rate to 72.7%. The latter rate is estimated based on the assumption that if we sent the mail follow-up to all nonrespondents to the 12-month online survey, they would respond at the same rate as the randomly selected subsample. Results from this additional mail assessment are described in Couper et al [15] and did not differ from data collected via the Internet or alter the results reported below [15].

Recruitment Methods

Two target populations were included in the study, and different recruitment approaches were used for each. The populations of

interest were (1) overweight health plan members generally and (2) overweight members with chronic illnesses for which weight management is a key part of treatment. The recruitment approaches were personal letters of invitation mailed to members' homes, for those with diabetes or coronary artery disease (CAD), and announcements about the study in mailed member newsletters and flyers posted at HMO facilities, for general members. An exception to this, which allowed a direct experimental comparison, is that in one HMO a sample of general adult members in one geographic region and a sample of members with hyperlipidemia also received personalized letters.

Personal Letters

Personal letters of invitation were sent to randomly sampled members in diabetes registries at all three health plans and in CAD registries at two plans. Letters were also sent to randomly selected overweight members of HMO 1 in a hyperlipidemia registry with no known CAD and to a random sample of the HMO 1 general adult membership. Since only members with BMI ≥ 30 (≥ 25 for those with diabetes or CAD) were eligible to participate in the study, BMI values recorded in the electronic medical record were examined for the two years prior to the beginning of study recruitment. Members whose most recently recorded BMI during this period was lower than the appropriate cutoff point were excluded from the sample. Members who had no BMI recorded during the two-year period were retained in the sample. Only members aged 18 years and over were considered. Table 1 gives the numbers of letters sent, by population and health plan.

The three HMOs employ consistent criteria for registry membership, using both the International Classification of Diseases, Ninth Revision (ICD-9) codes and specific pharmaceutical dispenses to identify patients. Patients in the diabetes registries were identified using either ICD-9 codes, specific pharmaceutical dispenses, or laboratory data indicating diabetes (two fasting blood sugars over 126 mg/dL or two random blood sugars over 200 mg/dL in the prior 12 months). Patients in the hyperlipidemia registries were identified by pharmaceutical dispenses and laboratory criteria. The two HMOs using CAD registries used the following criteria to identify patients with CAD. Patients had to have at least one of the following: (1) a hospital discharge (alive) with a principal or secondary diagnosis of acute myocardial infarction (AMI), percutaneous transluminal coronary angioplasty (PTCA), or coronary artery bypass graft (CABG), (2) a hospital discharge (alive) with a principal diagnosis of other acute or subacute ischemic heart disease, or (3) three or more outpatient visits with a diagnosis (principal or secondary) of CAD within a 36-month period.

The letters sent to potential participants were signed by physician leaders from the appropriate health plan. The letters gave instructions and pass codes for accessing the study website and a telephone contact at their plan. The letters described the study as an evaluation of online programs for helping people to lose weight and invited the addressee to participate in the study "if you are one of the thousands of people who say they want to lose weight." Members were told that to be eligible they must be overweight, a current member of the health plan, and have an email address and the ability to access the Internet at least once or twice a week.

Newsletters and Flyers

In HMOs 2 and 3, recruiting from the general membership was done primarily through general announcements in the quarterly member newsletters, although flyers describing the study were also posted and/or distributed at some local facilities. The newsletter announcements included a brief description of the study, gave plan-specific pass codes and instructions, and provided the name and telephone number of a local contact person. The newsletters were sent to all plan members in the region. Approximately 293000 and 121000 newsletters were distributed in the two regions, respectively. Based on state Behavioral Risk Factor Survey (BRFS) data on obesity rates, we estimate that 46827 and 30119 eligible adults received the newsletters, respectively, in HMO 2 and 3.

Program Descriptions

The HealthMedia Balance program has been described in detail elsewhere [1]. Briefly, it is a 6-week online self-help program that uses data from a baseline assessment to create an individually tailored weight management plan. Dimensions on which the program is tailored include health and medical history, prior weight loss efforts, intrinsic and extrinsic motivators for managing weight, perceived barriers to change, attitudes and stereotypes about overweight people and weight loss, social support systems, body image, nutritional habits, and physical activity. The program does not advocate a specific diet. Participants receive an initial program guide followed by newsletters containing tailored action plans 1 week, 3 weeks, and 6 weeks after enrollment.

While the Balance weight management program is the primary intervention in this study, we wanted to look at the potential impact of adding two additional Web-based, tailored interventions: Achieve and Nourish. Achieve is a goal-setting program delivered simultaneously with Balance that uses participant-reported performance data (ie, attributions for previous failure, motivation for continued performance, self-efficacy for continued performance) to determine follow-up questions and subsequent goals. The user may then adjust any goal according to his or her preference. Nourish is an online nutrition program that is very consistent with the Balance program. Nourish uses the same format for collecting information via an initial questionnaire to tailor nutrition advice to the specific needs and interests of the user. It consists of a guide, three sequential tailored newsletters, and email notifications delivered over an 8-week period. In this study, subjects received the Nourish program after first completing the Balance program.

Measures

This paper examines enrollment numbers and rates, engagement rates, and 12-month retention rates for the different recruitment approaches, target populations, and health plans. It also evaluates characteristics of members who did versus did not enroll in the study, those who did versus did not engage in the program at

different levels, and those who did versus did not complete the 12-month follow-up questionnaire.

Enrollment Rate

For members who received personal letters of invitation, the numerator of the enrollment rate is the number who enrolled in the study. The denominator is the total number of letters mailed minus the number of undeliverable letters. For members recruited through newsletters, the enrollment rate is the number who enrolled divided by the estimated total number of adult members estimated to be eligible, adjusted for the obesity rate in that state using Behavioral Risk Factor Survey data.

Levels of Program Engagement

Two measures of program engagement were calculated. To meet the criteria for initial engagement, participants needed to have viewed the initial electronic guide(s) relevant to their intervention. For example, participants in the Balance + Nourish intervention needed to have viewed the guides for both Balance and Nourish. The measure of ongoing engagement required, in addition, that participants view at least the initial follow-up electronic "newsletter" relevant to their intervention. These definitions were used because it was not possible in this study, given the data available, to examine engagement in more detail using continuous measures of program use.

12-month Retention Rate

The numerator of the retention rate is the number of participants who completed a 12-month follow-up survey, via either email or regular mail. The denominator is the total number of participants who enrolled.

Characteristics of Enrollees Versus Nonenrollees

Among the members who were sent letters of invitation, information is available in health plan records about both those who enrolled and those who did not. In order to protect the confidentiality of those who did not enroll, we obtained group-level de-identified aggregate data for enrollees and nonenrollees; thus, individual-level or multivariate analyses are not possible. Similar group data were also obtained for members who attempted to enroll in the study but did not meet eligibility criteria. Characteristics used to compare these groups included age distribution (10-year age bands), gender, age × gender distribution, smoking status (yes/no), BMI distribution, and proportions with certain specific medical conditions, such as diabetes.

To evaluate the potential influence of health status and prevalence of chronic conditions that could influence participation, we employed a pharmacy-based risk adjustment system called RxRisk to identify comorbidities across the cohort of members contacted. The RxRisk system is a revised and expanded version of the original chronic disease score (CDS) risk assessment instrument [17,18]. RxRisk is a clinically validated algorithm that estimates expected medical care cost for the patient for the next year based on chronic disease categories and prescription drug fills. Both a continuous variable, which we refer to as RxRisk, along with dichotomous variables associated with various comorbidities were created using pharmacy dispensings in the 12 months prior to the recruitment period. Given the non-normal distribution of the RxRisk score, we dichotomized the score at the median.

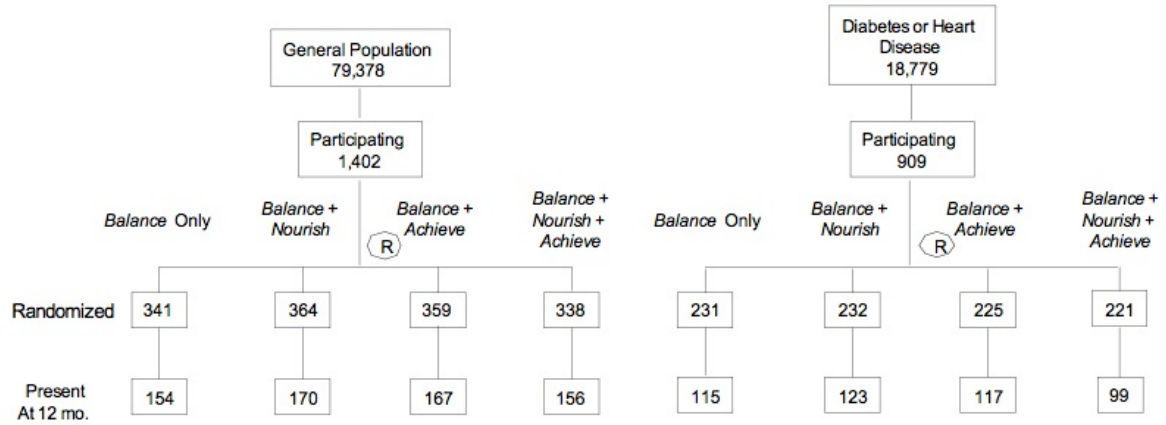
Characteristics of Participants With Respect to Program Engagement and Completion of the 12-Month Survey

Baseline measures on which participants were compared included demographic factors (age, gender, race/ethnicity), medical characteristics (BMI, presence of diabetes or CAD), and psychosocial factors (baseline motivation and self-efficacy for coping with stress). Motivation was measured by an item that asked respondents how motivated they currently were to manage their weight (on a scale of 0 to 10, 10 = highest motivation). Self-efficacy was measured with an item that asked respondents to rate their confidence in being able to manage their weight when stressed (1 = "not confident" and 5 = "extremely confident").

Analyses

The chi-square statistic was used to determine the significance of differences on various characteristics between members who participated in the study and those who chose not to participate. Because the data for these enrollment analyses were de-identified and aggregated, we used the chi-square statistic rather than a multivariate statistic that would require individual-level data. Among participants, multiple logistic regression was used to identify predictors of the level of program engagement and retention at 12 months. After examining independent variables and adjusting for any non-normality, associations between independent variables were examined to avoid multicollinearity and compromising the stability of the models. Due to strong associations between region and race/ethnicity, separate analyses were conducted, one with each of the associated independent variables. Additionally, diabetes was strongly associated with region, so diabetes and region were not used in the same regression model.

Figure 1. CONSORT diagram of participant retention



Results

A total of 2311 members enrolled, 909 of which had a chronic illness (see Figure 1, showing recruitment streams and attrition). Overall, 54% of participants were 60 years of age or older, 45% had a BMI of 35 or higher, over 40% had hypertension, 14% had diabetes, and 12% had CAD. Table 1 shows that outreach letters to members with a chronic illness elicited a higher enrollment rate (overall approximately 5%) than did similar letters to general members (2.4%). Among those in the general membership (without known disease), the direct experimental comparison revealed that those sent a personal letter participated at a higher rate (2.4%) than those receiving the newsletter announcement (1.7%).

Table 1. Participation rate among those sent personal letters

	Health Plan		
Population	**HMO 1**	**HMO 2**	**HMO 3**
Diabetes registry members	6.6% (97/1477)*	4.8% (219/4606)	0.7% (15/2054)
CAD registry members	–	4.6% (194/4220)	2.6% (92/3496)
Hyperlipidemia registry members	10.0% (292/2926)	–	–
General population in specific geographic region	2.4% (59/2432)	–	–

*number of enrollees / number of personal letters of invitation sent.

We also found differences in enrollment rates across the three HMO settings. The HMO having the lowest enrollment rate also had participated in a similar, widely advertised, Internet-based weight loss project previously and had a higher prevalence of African Americans than did the other two HMOs. The other health plans had not participated in the prior study, so it is difficult to draw conclusions about why these differences occurred. Table 2 summarizes results from analyses on demographic and medical characteristics of those members who enrolled and those who declined. Those who enrolled were younger, were more likely to be nonsmokers and female, and were more likely to have lower RxRisk scores (less disease risk or burden, $P < .001$).

Table 2. Characteristics of enrollees and decliners

Characteristic	Enrollees (%)	Decliners (%)	*P* value
Age			
< 60 years	53.5	40.0	< .001
> 60 years	46.5	60.0	
Gender			
Male	46.7	58.1	< .001
Female	53.3	41.9	
Smoking Status			
Current smoker	5.7	12.2	< .001
Nonsmoker	94.3	87.8	
RxRisk (estimated medical costs)			
< US $3000	55.6	44.3	< .001
> US $3000	44.4	55.7	

Engagement

Program engagement varied widely by intervention (Table 3). The vast majority of participants viewed the initial Balance guide, and approximately half met the criteria for ongoing Balance engagement. As can be seen, substantially fewer participants met criteria for engagement in the other interventions. This might be expected since these criteria were more stringent and required more of participants (eg, those in the Balance + Achieve intervention had to meet requirements for both the Balance and Achieve programs). Both initial and ongoing engagement rates were especially low for the Nourish intervention, likely because this program was only initiated after the initial intervention period for Balance users.

Table 3. Percent of patients achieving different levels of program engagement, by intervention

Intervention	Initial Engagement* (%)	Ongoing Engagement† (%)
Balance only (n = 572)	90.9	49.0
Balance + Achieve (n = 584)	62.2	25.3
Balance + Nourish (n = 596)	19.1	8.1
Balance + Achieve + Nourish (n = 559)	13.4	5.7

*Initial Engagement = Viewed initial electronic guides appropriate to that intervention

† Ongoing Engagement = Initial engagement plus viewed at least initial electronic follow-up newsletters appropriate to intervention (or for Achieve, set at least initial goal)

As can be seen in Table 4, the intervention was the primary factor associated with initial engagement (the reference category for treatment effects was participants not randomized to a given intervention). All of the significant treatment effects and interactions with treatment indicated that the basic Balance intervention produced substantially higher initial engagement rates than the others. The only other variable to reach significance was gender, with females being significantly more likely than males to view initial guides ($P < .001$).

Analyses of ongoing engagement revealed a number of significant associations (lower portion of Table 4). All of the factors significant for initial engagement were also significant predictors of ongoing engagement. In addition, older participants ($P < .001$) and those with higher baseline motivation levels ($P = .04$) were more likely to demonstrate ongoing engagement. In contrast, African Americans ($P = .03$) and those with higher baseline self-efficacy scores ($P = .003$) were less likely to be engaged with their program on an ongoing basis.

Table 4. Results of logistic regression to predict engagement

Factor	Odds Ratio (CI)	Beta	SE	*P* value
Predictors of Initial Engagement (N = 2276)				
Nourish (vs non)	0.02 (0.02-0.03)	−3.78	0.18	< .001
Achieve (vs non)	0.17 (0.12-0.23)	−1.79	0.17	< .001
Nourish × Achieve (Balance only)	3.97 (2.49-6.32)	1.38	0.24	< .001
Baseline BMI	0.99 (0.97-1.00)	−0.02	0.01	.10
Age	1.01 (1.00-1.02)	0.01	0.01	.27
Female	1.68 (1.27-2.22)	0.52	0.14	< .001
Ethnicity (see below)				.10
Diabetes diagnosis (vs non)	0.94 (0.73-1.20)	−0.07	0.13	.60
CAD diagnosis (vs non)	1.16 (0.83-1.60)	0.14	0.17	.39
Self-efficacy	0.95 (0.85-1.06)	−0.06	0.06	.32
Motivation	1.03 (0.97-1.10)	0.03	0.03	.27
Predictors of Ongoing Engagement (N = 2276)				
Nourish (vs non)	0.09 (0.06-0.12)	−2.46	0.18	< .001
Achieve (vs non)	0.34 (0.27-0.44)	−1.07	0.13	< .001
Nourish × Achieve (Balance only)	2.04 (1.19-3.48)	0.71	0.27	.009
Baseline BMI	0.99 (0.97-1.01)	−0.01	0.01	.44
Age	1.02 (1.01-1.03)	0.02	0.01	.001
Female	1.50 (1.12-2.01)	0.41	0.15	.006
Ethnicity				.04
White (vs non)	1.18 (0.92-1.51)	0.16	0.13	.20
Black / African American (vs non)	0.68 (0.47-0.97)	−0.39	0.18	.03
Hispanic/Latino	0.88 (0.56-1.38)	−0.13	0.23	.57
Diabetes diagnosis (vs non)	0.97 (0.76-1.25)	−0.03	0.13	.81
CAD diagnosis (vs non)	1.01 (0.73-1.40)	0.01	0.17	.97
Self-efficacy	0.84 (0.75-0.95)	−0.17	0.06	.003
Motivation	1.07 (1.00-1.14)	0.07	0.03	.04

Retention

Approximately 48% of initial participants provided information at 12-month follow-up through either online or mailed surveys. Logistic regression analyses on characteristics of respondents and nonrespondents to this follow-up revealed a few significant factors (Table 5). Younger enrollees and those who had higher baseline levels of self-efficacy were less likely to participate in the follow-up. However, there were no significant effects of intervention type, BMI, gender, baseline motivation level, ethnicity, or presence or absence of either diabetes or CAD.

Table 5. Results of the multiple regression to predict retention at 12 months

Factor	Odds Ratio (CI)	Beta	SE	*P* value
Nourish (vs non)	1.11 (0.86-1.37)	0.08	0.12	.50
Achieve (vs non)	1.08 (0.83-1.33)	0.05	0.12	.67
Nourish × Achieve (vs non)	0.87 (0.59-1.15)	−0.19	0.17	.26
Baseline BMI	0.99 (0.98-1.01)	−0.01	0.01	.22
Age	1.01 (1.00-1.02)	0.01	0.004	< .001
Ethnicity*	1.00-1.19	−0.05 to 0.17	0.09-0.16	.32
Gender	0.94 (0.71-1.08)	−0.13	0.11	.23
Baseline self-efficacy	0.92 (0.84-0.99)	−0.10	0.04	.03
Baseline motivation	1.01 (0.96-1.06)	0.01	0.02	.75
Diabetes diagnosis (vs non)	0.99 (0.81-1.18)	−0.02	0.10	.82
CAD diagnosis (vs non)	0.87 (0.66-1.09)	−0.17	0.13	.19

*Ethnicity involved three separate contrasts: Hispanic vs Other, African American vs Other, and Non-Hispanic White vs Other.

Discussion

Many HMO members are willing to participate in Internet-based weight management programs. Although the overall participation rate was not high in an absolute sense, eHealth programs may be an efficient way of delivering health promotion services to a large number of members. This would especially be so if these programs attract representative or high-risk participants. Representativeness analyses are essential to evaluate the public health impact of eHealth programs [5]. Our results were mixed on this issue: although we did attract many older patients with chronic illness, the enrollment analyses suggest that in general this program did not attract those at highest risk (smokers, older adults, etc) at the same rate as those at lower risk. However, among members in HMO 1 receiving letters of invitation, patients with CAD or diabetes were substantially more likely to participate than members without known chronic illness (10% and 7%, respectively, vs 2.4%).

Congruent with other computer-mediated programs, it appears that sending personal letters from health professionals is an effective method of enhancing enrollment rate [19-21]. Newsletter articles and flyers, in contrast, are a low-cost alternative but may recruit a considerably smaller proportion of the target audience. Although our evaluations of the type of recruitment and disease status were not randomized and were not orthogonal comparisons (with the exception of the comparisons within HMO 1), the results are fairly clear that personalized mailings increase enrollment rates, and members with a diagnosed chronic disease are more likely to participate than are other members.

Our engagement analyses revealed that adding components to a basic Internet-based intervention program can create adherence challenges. Although almost all participants viewed the initial Balance materials, far fewer viewed the other electronic guides in the Achieve and Nourish interventions. Intervention assignments were the factors most strongly associated with both initial and ongoing engagement; but, in addition, females were more likely to become and remain engaged than males. Additional demographic and motivational factors also predicted ongoing engagement (but not initial engagement). From an adherence perspective, it also appeared more successful to introduce additional treatment components during initial weight management stages (as in Achieve) than to wait until later (as in Nourish). Future studies should evaluate the use of the dichotomous engagement criteria used in this study compared to more sophisticated, automated engagement measures, such as patterns of log-ins over time.

As in many eHealth studies, there was substantial attrition by the time of the 12-month follow-up. Our attrition analyses suggest that those who declined to participate in the follow-up were generally representative of those who continued participation, the primary exceptions being baseline level of self-efficacy and age. Other recent eHealth research has found that dropping out of assessment is different from dropping out of intervention, and that those who drop out of eHealth programs may benefit just as much as those who do not [15], in contrast to typical findings in clinician- or educator-delivered programs. In terms of health disparities and health impact, it was encouraging that those who were older, weighed more, had diagnosed disease, and were members of racial/ethnic minority groups were equally or more likely to participate in follow-up assessments.

This study had both strengths and limitations. Strengths include the study of an eHealth program that proved efficacious in prior research [1]; analyses of a large-scale implementation of this program in three different “real-world” health care settings; the relatively comprehensive analyses of characteristics and the representativeness of those who enrolled versus declined, made possible by the electronic health record systems of the participating organizations; the inclusion of reach, engagement, and retention analyses; and the study of different recruitment methods. Limitations include the inability to conduct multivariable analyses of reach due to the Health Insurance Portability and Accountability Act (HIPAA) and privacy issues, the absence of information on some important factors such as

health literacy, and the confounding of some issues such as diagnosed disease status and recruitment method.

Future research should compare the reach of Internet-based and other modalities of health promotion and investigate methods to enhance ongoing engagement and retention, which may be particular challenges for eHealth programs.

Acknowledgments

This paper was supported by a grant from the Robert Wood Johnson foundation to HealthMedia, Inc. (Kevin Wildenhaus, PhD, principal investigator) under their eHealth initiative.

Conflicts of Interest

Drs. Strecher and Wildenhaus are associated with HealthMedia, Inc., which developed and has proprietary interests in the three weight management programs described in the paper. Therefore, all analyses for this article were supervised and performed by personnel not part of HealthMedia.

Multimedia Appendix

Representativeness in eHealth Programs: Factors Related to Recruitment, Participation, and Retention - Poster (ppt) [PPT file (MS Powerpoint), 5 MB - jmir_v9i2e11_app1.ppt]

References

1. 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]
2. 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]
3. Winett RA, Tate DF, Anderson ES, Wojcik JR, Winett SG. Long-term weight gain prevention: a theoretically based Internet approach. Prev Med 2005 Aug;41(2):629-641. [Medline: 15917062] [doi: 10.1016/j.ypmed.2004.12.005]
4. Couper MP. Issues of Representation in eHealth Research (with a Focus on Web Surveys). Am J Prev Med 2007 May;32(5 Suppl):S83-S89. [Medline: 17466823] [doi: 10.1016/j.amepre.2007.01.017]
5. Glasgow RE. eHealth Evaluation and Dissemination Research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [Medline: 17466816] [doi: 10.1016/j.amepre.2007.01.023]
6. 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]
7. Brug J, Oenema A, Campbell M. Past, present, and future of computer-tailored nutrition education. Am J Clin Nutr 2003 Apr;77(4 Suppl):1028S-1034S [FREE Full text] [Medline: 12663313]
8. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, Mcghee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [Medline: 15631964] [doi: 10.2196/jmir.6.4.e40]
9. Williamson DA, Martin PD, White MA, Newton R, Walden H, York-Crowe E, et al. Efficacy of an internet-based behavioral weight loss program for overweight adolescent African-American girls. Eat Weight Disord 2005 Sep;10(3):193-203. [Medline: 16277142]
10. Koo M, Skinner H. Challenges of internet recruitment: a case study with disappointing results. J Med Internet Res 2005;7(1):e6 [FREE Full text] [Medline: 15829478] [doi: 10.2196/jmir.7.1.e6]
11. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annu Rev Public Health 2006;27(1):1-28. [Medline: 16533107] [doi: 10.1146/annurev.publhealth.27.021405.102113]
12. Vollmer WM, Svetkey LP, Appel LJ, Obarzanek E, Reams P, Kennedy B, et al. Recruitment and retention of minority participants in the DASH controlled feeding trial. DASH Collaborative Research Group. Dietary Approaches to Stop Hypertension. Ethn Dis 1998;8(2):198-208. [Medline: 9681285]
13. Im EO, Chee W. Methodological issues in the recruitment of ethnic minority subjects to research via the Internet: a discussion paper. Int J Nurs Stud 2005 Nov;42(8):923-929. [Medline: 16210030] [doi: 10.1016/j.ijnurstu.2005.01.002]
14. Glasgow RE, Boles SM, Mckay HG, Feil EG, Barrera M. The D-Net diabetes self-management program: long-term implementation, outcomes, and generalization results. Prev Med 2003 Apr;36(4):410-419. [Medline: 12649049] [doi: 10.1016/S0091-7435(02)00056-7]
15. Couper MP, Peytchev A, Rothert KLittle RJA. Combining information from multiple modes to evaluate and reduce nonresponse bias. In: Proceedings of the American Statistical Association, Survey Research Methods Section. Alexandria, VA: American Statistical Association; 2005.

16. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [Medline: 15829473] [doi: 10.2196/jmir.7.1.e11]
17. Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. Med Care 1995 Aug;33(8):783-795. [Medline: 7637401] [doi: 10.1097/00005650-199508000-00004]
18. Fishman PA, Goodman MJ, Hornbrook MC, Meenan RT, Bachman DJ, O'keeffe Rosetti MC. Risk adjustment using automated ambulatory pharmacy data: the RxRisk model. Med Care 2003 Jan;41(1):84-99. [Medline: 12544546]
19. Glasgow RE, Strycker LA, King DK, Toobert DJ, Rahm AK, Jex M, et al. Robustness of a computer-assisted diabetes self-management intervention across patient characteristics, healthcare settings, and intervention staff. Am J Manag Care 2006 Mar;12(3):137-145 [FREE Full text] [Medline: 16524346]
20. Glasgow RE, Nutting PA, King DK, Nelson CC, Cutter G, Gaglio B, et al. A practical randomized trial to improve diabetes care. J Gen Intern Med 2004 Dec;19(12):1167-1174 [FREE Full text] [Medline: 15610326] [doi: 10.1111/j.1525-1497.2004.30425.x]
21. Glasgow RE, Nutting PA, Toobert DJ, King DK, Strycker LA, Jex M, et al. Effects of a brief computer-assisted diabetes self-management intervention on dietary, biological and quality-of-life outcomes. Chronic Illn 2006 Mar;2(1):27-38. [Medline: 17175680] [doi: 10.1179/174592006X93851]

Abbreviations

BMI: body mass index
CAD: coronary artery disease
HMO: health maintenance organization

submitted 09.08.06; peer-reviewed by H Christensen; comments to author 21.12.06; revised version received 19.01.07; accepted 21.03.07; published 09.05.07.

Please cite as:
Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, Couper MP, Green B, Wildenhaus K
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
URL: http://www.jmir.org/2007/2/e11/
doi:10.2196/jmir.9.2.e11
PMID:17513282

Original Paper

How Complementary and Alternative Medicine Practitioners Use PubMed

John Willinsky[1], MEd, PhD; Mia Quint-Rapoport[2], MA

[1]University of British Columbia, Vancouver, BC, Canada

[2]University of Toronto, Toronto, ON, Canada

Corresponding Author:
John Willinsky, MEd, PhD
Department of Language and Literacy Education
University of British Columbia
Vancouver, BC V6T 1Z4
Canada
Phone: +1 604 822 3154
Fax: +1 604 822 3154
Email: john.willinsky@ubc.ca

Abstract

Background: PubMed is the largest bibliographic index in the life sciences. It is freely available online and is used by professionals and the public to learn more about medical research. While primarily intended to serve researchers, PubMed provides an array of tools and services that can help a wider readership in the location, comprehension, evaluation, and utilization of medical research.

Objective: This study sought to establish the potential contributions made by a range of PubMed tools and services to the use of the database by complementary and alternative medicine practitioners.

Methods: In this study, 10 chiropractors, 7 registered massage therapists, and a homeopath (N = 18), 11 with prior research training and 7 without, were taken through a 2-hour introductory session with PubMed. The 10 PubMed tools and services considered in this study can be divided into three functions: (1) information retrieval (Boolean Search, Limits, Related Articles, Author Links, MeSH), (2) information access (Publisher Link, LinkOut, Bookshelf), and (3) information management (History, Send To, Email Alert). Participants were introduced to between six and 10 of these tools and services. The participants were asked to provide feedback on the value of each tool or service in terms of their information needs, which was ranked as positive, positive with emphasis, negative, or indifferent.

Results: The participants in this study expressed an interest in the three types of PubMed tools and services (information retrieval, access, and management), with less well-regarded tools including MeSH Database and Bookshelf. In terms of their comprehension of the research, the tools and services led the participants to reflect on their understanding as well as their critical reading and use of the research. There was universal support among the participants for greater access to complete articles, beyond the approximately 15% that are currently open access. The abstracts provided by PubMed were felt to be necessary in selecting literature to read but entirely inadequate for both evaluating and learning from the research. Thus, the restrictions and fees the participants faced in accessing full-text articles were points of frustration.

Conclusions: The study found strong indications of PubMed's potential value in the professional development of these complementary and alternative medicine practitioners in terms of engaging with and understanding research. It provides support for the various initiatives intended to increase access, including a recommendation that the National Library of Medicine tap into the published research that is being archived by authors in institutional archives and through other websites.

(J Med Internet Res 2007;9(2):e19) doi:10.2196/jmir.9.2.e19

KEYWORDS

PubMed; research dissemination; complementary and alternative medicine; open access; professional development; information retrieval; information management; literacy

Introduction

This study considers the use of PubMed (www.pubmed.gov) by practitioners of complementary and alternative medicine (CAM). In 2001, Americans invested US $50 billion in this form of health care, according to John Weeks, editor of *The Integrator*, a newsletter tracking "the business of alternative medicine," with 40% of American adults turning at some point to alternative medicine [1]. While some in the medical research community have argued that "there is no alternative medicine" as "there is only scientifically proven, evidence-based medicine supported by solid data, or unproven medicine for which there is no scientific evidence" [2], the US National Center for Complementary and Alternative Medicine (one of the National Institutes of Health) sponsors research in this field and serves as a clearinghouse for clinical trials and research summaries.

Within the scope of its standing as the largest bibliographic index on the life sciences, PubMed (containing Medline, formerly known as Index Medicus, as a subset) indexes medical research articles and includes material useful for physicians and pharmacists, as well as traditional Chinese medicine practitioners [3,4]. A study of nurses, health care assistants, midwives, and health visitors (registered nurses or midwives who promote mental, physical, and social well-being in the community) has shown that, among those who had sufficient access to online sources of health information, the access to journal literature was felt to have led to improved patient care [5].

A prerequisite for the successful pursuit of health information is what Norman and Skinner identify as eHealth literacy, or the skills that enable people "to seek, find, understand, and appraise health information" [6]. This study focuses on how, among a group of CAM practitioners, those abilities are supported by an array of PubMed tools and services. Up to this point, investigators have recognized that users of online health services are not a monolithic group. Studies have been conducted on the reading levels demanded of cancer patients [7], the overwhelming quantity of information available for first-time mothers [8], the limited search skills of college students studying health [9], the growing use of electronic resources by nurses with limited search skills [10,11], and the growing confidence among doctors in their search strategies, although few use this skill on a real-time basis with patients [12].

The variety of approaches that these studies apply testifies to the idea that reading health literature online is situation specific and calls for context, and that investigations into eHealth literacy are fruitful when specific populations are delineated. Information retrieval studies have been done on citation retrieval [13] and literature searching [14] of PubMed, as well as on the use of the medical subject heading (MeSH) and Limits services [15]. Work has also been done in advising users how to best conduct CAM searches [16,17] with at least one study finding that PubMed indexing is not yet adequate for this field of medicine, especially with CAM clinical trials [17].

This study is intended to complement the more common focus among information science researchers on users' information retrieval strategies with indexing services such as PubMed [18,19]. It addresses the value of PubMed's specific tools and services for CAM practitioners, as these tools are introduced to them in the context of their own interests in the life sciences and their interest in learning about the use of PubMed. In focusing on the value and role of the tools and services, we are drawing on literacy and learning research in the subject domains, or "domain knowledge" as Alexander and others refer to it [20]. Alexander has established the degree to which a combination of personal interest and background knowledge contribute to the degree of learning in domain areas such as biology and physics, and she holds, in particular, that "one's knowledge base is a scaffold that supports the construction of all future learning" [21].

In this sense, PubMed's provision of tools and services (such as Related Articles and Bookshelf) can be seen to help users easily and immediately augment, in a limited way, shortcomings in their knowledge base. The tools and services compensate for shortcomings not only in searching for articles but also when comprehending, evaluating, and utilizing abstracts and full-text articles. The tools and services (such as Full-Text Access, Email Alert, Send To) also appear to be poised to support the user's interest in and engagement with the indexed material (which is the other critical factor Alexander has identified for domain learning). By introducing CAM practitioners to the tools and services, one by one, and gaining from them a sense of the perceived value and contribution of each, this study seeks to determine the potential contribution of these tools and services to the knowledge-base scaffolding and the personal interest of health care workers who lie outside of PubMed's originally intended audience of researchers and physicians. This work has obvious implications for the design of the tools, in PubMed's effort to serve as broad an audience of health care practitioners as possible, and as a way, in turn, of extending the impact and benefits of life science research.

Methods

Sample

The participants in this study were drawn from among those practising CAM. They were recruited from three sources: (1) participants 1-5 were drawn from among students taking an online professional development course at a community college on research literacy for the health professions, (2) participants 6-9 were recruited from the Massage Therapy Association of British Columbia, and (3) participants 10-18 were drawn from the Canadian Memorial Chiropractic College. The eight women and 10 men who agreed to participate in the study ranged in age from their 20s to their 50s. Of the 18 participants, 10 were pursuing or possessed a chiropractic degree, seven were registered massage therapists, and one practised homeopathy. A total of 13 participants had a university degree, with one holding a master's degree and another, a PhD. In terms of background with online research, participants 1-5 were taking a course on research literacy, while participants 13-18 had been introduced to the topic in a session with a librarian, with some one-on-one follow-up.

Design

Participants were offered an opportunity to learn how to use PubMed in one-on-one sessions with the researcher (MQ-R) in exchange for helping the researchers understand the value of PubMed's various features to users such as themselves. The 10 PubMed tools and services considered in this study (Table 1) can be divided into three functions: information retrieval (Boolean Search, Limits, Related Articles, Author Links, MeSH), information access (Publisher Link, LinkOut, Bookshelf), and information management (History, Send To, Email Alert). The sessions with the participants took place between April 2004 and December 2006 and averaged 2 hours in length. Sessions held with participants 1-9 were conducted over the phone, with both participant and researcher accessing PubMed through the Internet, while sessions with participants 10-18 were conducted face-to-face with a single computer connected to the Internet, which the participant operated. Participants were introduced to between six and 10 PubMed tools. The participants in the phone sessions covered 7.1 items on average compared to 7.7 items covered by participants in the face-to-face sessions.

Prior to their PubMed training session, the researcher asked participants by email to identify a health science topic of interest to them. The researcher then found a suitable article based on the interest of the 14 participants who responded and sent it to them in advance of the PubMed session (Multimedia Appendix 1). The participants were asked to read the paper prior to the session. The starting point for each session, after introductions and preliminary questions about the participant's background, was an initial search of PubMed on a topic of interest. Participants were then guided through a range of six to 10 PubMed tools and services depending on the pace of the 2-hour session and where the interests of the participants led (Multimedia Appendix 2). For each tool and service, they were introduced to how the feature worked, using concepts and papers of interest to them, and were asked to provide their thoughts on its potential value for their use of PubMed.

Table 1. PubMed tools and services introduced to participants

Tool/Service	Function
Boolean Search	Search terms can be entered using Boolean operators AND, OR, NOT.
Search Limits	Allows the user to limit the search by language; type of journal; gender; human or animal subjects; age; type of article; date; inclusion into PubMed.
Related Articles	This appears as a button/link with each citation and brings up a list of matching articles based on the original article.
Full-Text Access	This is provided through the publisher's link on the abstract page and on the LinkOut page, with publisher, aggregator, PubMed Central, and/or subscribing libraries providing open access for roughly 15% of articles.
Send To	Send To allows users to email citations, convert citations to a text file, or save citations for editing and further action.
Search History	This tracks terms, time, and number of results in searches, while allowing Boolean searches combining different searches from the Search History.
Bookshelf	This provides links to relevant passages in a set of full-text life science books that range from basic science to clinical practices.
MeSH Vocab	MeSH is a controlled vocabulary database for ascertaining the most commonly used terms in PubMed.
Email Alert	The National Center for Biological Information (NCBI) of the National Library of Medicine (NLM) offers email notification of new work in specified topics.
Author Links	Authors' names have links leading to a listing of all of their work in PubMed.

Analysis

The sessions were transcribed (Multimedia Appendix 3), and the two researchers created a summary table of the results by rating the responses for each tool as (+) positive ("that's valuable"), (++) positive with emphasis ("I'd really use that"), or (−) negative or indifferent ("that makes sense"). Because of the session's 2-hour time constraint and the semistructured handling of the sessions, not every tool was covered with every participant. As a follow-up to the PubMed session, the researcher emailed the participants 2 weeks later, asking them to describe their current use, if any, of PubMed and other Internet resources.

Results

What follows is a detailed qualitative analysis of the participants' responses to each of the 10 PubMed tools and services covered in this study, with the aim of better understanding how such aspects of PubMed might be seen as contributing to its use by the participants. Differences in the responses among the participants (on the basis of training or university degree) was not the focus of the study, although, as might be expected, those with prior training and with university degrees tended to value PubMed tools and services more highly. Participants with prior training in the use of online research literature thought that, on average, 6.0 out of the 7.5 tools and services that were introduced to them would be valuable in their use of PubMed, while those without such training responded positively to 4.6 of the 7.1 items introduced to them (Table 2). Similarly, those with a university degree responded positively to 5.7 out of 7.4 items, compared to those without a university degree, who gave a positive evaluation to 4.8 out of 7.4 items.

Table 2. Value of PubMed tools and services for participants (N = 18)

Participants			PubMed Tools and Services[‡]										
Code[†]	University Degree	Prior Online Research Training	Boolean Search	Search Limits	Related Articles	Full-Text Access	Send To	Search History	Bookshelf	MeSH Vocab	Email Alert	Author Links	Further Use of PubMed
MF1	yes	yes	+[*]	+[*]	+	++	+	++[*]	−	−[*]	+[*]		yes
CM2	yes	yes	++	+	++	+	+[*]		−				yes
HF3	no	yes	++[*]	+[*]	++[*]	+	+	++	−			+	no
MF4	yes	yes	+[*]	+[*]	+[*]	+[*]	++		++				
MF5	yes	yes	+[*]	+	+	++			++	+			
MM6	no	no	+	++	++	++[*]	−		+	−		−	
MF7	no	no	++	−	++	+	+		++				yes
MF8	no	no	−	−	+	++	−		−	+		−	yes
MM9	yes	no	+	−	−	+	−			++		−	no
CF10	no	no	++	++	−	+	−	+		−			yes
CM11	yes	no	++	−	−	++	−	++		+	++	+	yes
CF12	yes	no	+	+	−	++	++			++			yes
CM13	yes	yes	+	−	+	+	−	+	−		+	+	no
CM14	yes	yes	+	−	+[*]	+	+[*]	++[*]	−		+		
CM15	yes	yes	−	++	++	++	++	+[*]	+	−			yes
CM16	yes	yes	+[*]	+	++	+[*]	++	+	−				
CM17	yes	yes	+[*]	+[*]	+	++[*]	++	+	−	−[*]	+		
CM18	yes	yes	−	++	−[*]	+[*]	+	−	+				

[*]Prior familiarity with tool or service.

[†]First letter of code: M=registered massage therapist; C=chiropractor; H=homeopath. Second letter of code: M=male; F=female.

[‡] +: positive evaluation of tool or service; ++: additional positive emphasis; −: negative or indifferent response to tool or service

Blank cells indicate that tool or service was not covered in the participant's session or participant did not respond to email follow-up.

Boolean Search

The basic starting point with PubMed, as with any online index or database, is the search, which typically allows the combining of search terms with AND, OR, or NOT for more accurate searching. These Boolean operators were familiar to a number of participants, but were a little vague for others: "I might have heard of it" (CF12). Learning about them made a considerable impression on a number of participants:

> *I will use the Boolean search methods even when using the different features of the database. This is incredibly important to know, use, and understand. It's like the building blocks of searches. I've got to practise that one afterwards.* [CM2]

At least one participant recognized that the search results were profoundly influenced by the terms of the search:

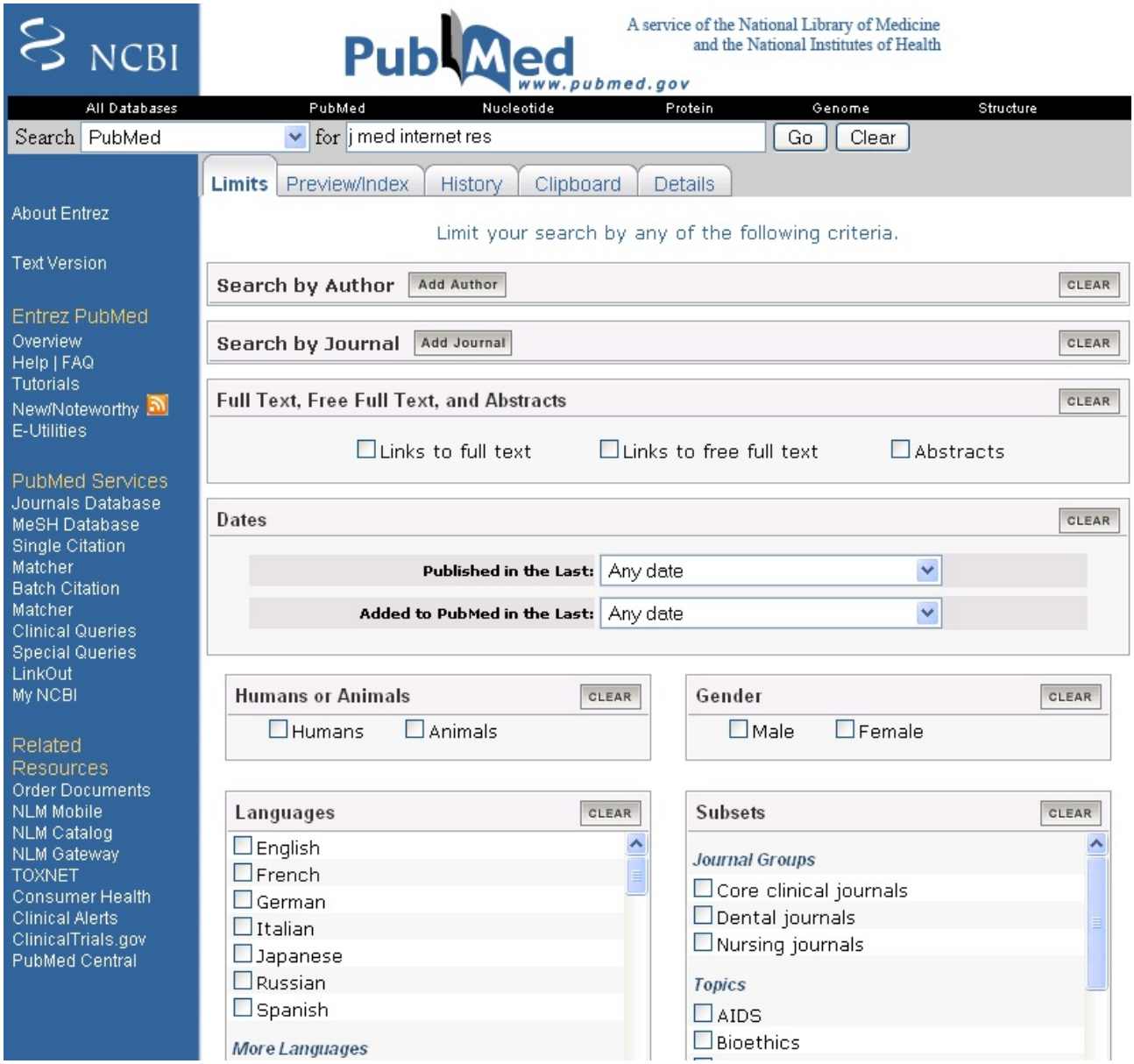
> *The difficulties is [sic] the prejudice you take to your search question. Maybe you missed something because you describe[d] it badly.* [CM11]

Search Limits

A further search strategy that the participants found useful was the ability to apply Limits to the searches, including dates, author, links to free full-text articles, language, gender, human subjects, type of article, and so on (Figure 1): "I have used Limits with dates mainly...and if I am searching for a particular author" (MF1). Limits seemed useful for participants for finding specific information, such as when they needed to answer a particular question: "I think I would use this one [Limits] to look for something very specific, but if I wanted to browse around I would use Related Articles" (MF5), and "I think I would use the Limits if I have a problem patient who comes at me with a Google paper, like if she was a 45-year-old female who found something and said, 'Why don't I try this?'" (CF10). The participants found Limits useful because it allowed them to be selective about the information they would find, such as limiting their searches specifically to include only core clinical journals. This is because "they [core clinical journals] are well recognized in the medical community" (MF7), and, as another participant explains, "Because sometimes I don't want to

know...what a really alternative journal says, I want to read the larger stuff" (CF12).

Figure 1. Limits tab on PubMed



Related Articles

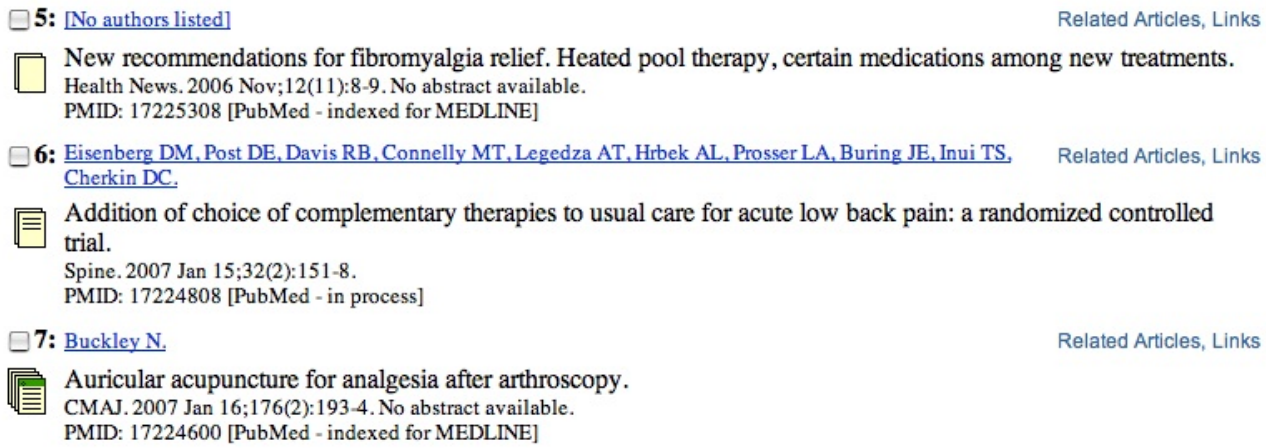
The ability to identify one very promising article and then find all of the related articles by clicking on a hyperlink (Figure 2) was highly valued by five participants: "I really like Related Articles; I saw it in the other database, and it is useful because it brings me to things that I would probably not have found otherwise, outside my own searching" (CM2). However, participants also intuited how Related Articles could be used in relation to Boolean searches to achieve a very efficient focus on the desired research: "These look really great.… I could spend hours just seeing what comes up from this feature after starting from one search" (HF3), and "This list [based on Related Articles] makes more sense to me than the original list that came up" (MF4). It did seem to a number of the participants that "Related Articles worked perfectly" (MF7). Yet, Related Articles did not work for everyone, as one chiropractor commented, "What [PubMed] thinks is similar is not what I think is similar" (CM18).

Full-Text Access

PubMed provides access to the complete articles that it indexes through a publisher's link on the article's abstract page (see Figure 2), as well as a link to PubMed Central if the journal's contents have been placed in this open-access repository. In addition, LinkOut provides links to sites where the article can be accessed, including sites of the publisher, journal aggregators such as EBSCO and ProQuest, and university libraries that hold print and/or electronic editions of the article (which are accessible to members of those libraries). Many of the publishers and the aggregators provide nonsubscribers the ability to purchase access. For example, Elsevier's Science Direct changes US $30 to purchase an article in any of its 2200 journals. In addition, PubMed clearly identifies that a small proportion (roughly 15%) of the literature has been made "open access"

by its publisher, either immediately upon publication or after a certain period ("moving wall" model), typically from 6 to 24 months following publication.

Figure 2. Partial results of a PubMed search on "complementary medicine," showing icons (from top to bottom) for "no abstract," "abstract only," and "open access to full-text article" (although no abstract is available) and hyperlinks to "Related Articles" and "Links" (to the full text on publisher and/or library websites)



While the article's abstract was typically included in PubMed's indexing of an article, it was judged insufficient by participants as a way of both understanding and evaluating the research, whether they had had very little experience with the research or were well oriented to their field's literature. As one first-time PubMed user stated as she looked at an abstract, "[I] definitely need to read the article. It [abstract] does not tell us enough" (MF8). This response is consistent with previous studies on the shortcomings of abstracts [22-24].

For most participants, the abstract was necessary as a starting point but was insufficient as source of information on its own. It served as a screen, as it would for a researcher:

> *So an abstract is really useful. If I felt the abstract was interesting, I would definitely need the full text to actually find the article useful to my work. An abstract only provides enough information to pass the first screening test.* [CM2]

This chiropractor initially felt, on first looking through the list of PubMed search results, that "This information is not what I am looking for; it is too technical, too specialized." However, by the time he had had the chance to download and look at an open-access article, he was expressing regrets: "This is incredible; it is so easy; I wish I could just finish reading the article right now" (CM2). Frustration over not being able to view full-text articles was a common theme: "There was an article that I was interested in looking at, and it said that it would not be available for about a year or 6 months" (CM14). "The only problem," as a registered massage therapist put it, "is that the ones that are free are not the ones I was wanting" (MF8).

For some, visiting a research library was not an option, and the cost of purchasing a single article posed a serious obstacle: "It deters me, big time, having to pay for articles" (MF8), and "Oh, you can order papers; I don't have money to order papers" (CM14). For others, using a credit card online was an issue: "I don't think that that's a safe measure, with my credit card online" (MF5). Yet the interest in seeing an article was sufficient that a number of participants were prepared to give serious consideration to paying to download it: "It is incredibly important to have access to the full text of the article. I would even pay to access a full-text article" (CM2); "I think I'll look for mostly free articles so I could actually read them. If I was desperate to find information and couldn't find anything, then I might actually buy it" (HF3), and "depending on how valuable the article is I might be willing to do that" (MF7).

For one chiropractor, who administered a professional chiropractic organization, having access to the full-text article was a must. He felt that the full text offered a more accurate picture of the research and provided the language he needed to use for his own work, which was to lobby the Canadian government to further integrate chiropractic into standard public health policies. As he explained,

> *No one who has ever done research has ever done an abstract that exactly covers the detail of what's done inside. So when you actually read the research paper you see more…. The abstract…never provides that robust meaty quote that I need.* [CM11]

He went on to give an example from the *New England Journal of Medicine* in which the results reported in the body of the article placed the chiropractor's treatment of childhood asthma in a far better light than the abstract did.

Another chiropractor conveyed the frustration of coming across a very good source of research, only to find access to the resource restricted. In this case, it was a systematic review conducted by the Cochrane Collaboration, which are not freely available:

> *It says [reading from the abstract] "evidence which is based on the Cochrane review of the subject." This review is like topnotch.... It's called a Cochrane collaboration, and we don't have free access to that.... It would be phenomenal if we had access to that.... It's all based on a strict protocol, like how many people were in the study.... If you want to know the vast information about a subject, you should go to Cochrane reviews.* [CM14]

Even in the case of a chiropractor who felt that "often times I get most of the information I need from the abstract" (CM17) and thus did not need the full text, there was a recanting after a few moments:

> *Well, that's not true. I do [need access to the full text]. Often times you are stuck because they tell you "in the article we describe" such and such. And then you do everything to get it. And we have different ways to get articles.* [CM17]

A registered massage therapist noted how much she appreciated it when the researcher sent, prior to her session, the full text of an article of interest:

> *I especially valued that it was not just a summary.... It is very relevant to my work as a therapist, and as an instructor. [I] also found the references very helpful.* [MF7]

And finally, a registered massage therapist who had recently obtained her MSc from a Canadian university, and thus had lost her ready access to the scholarly literature upon graduation, had learned about the issue of access in terms of how libraries were signing a new type of agreement with the journal publishers, which carefully controlled access:

> *The contracts that the academic institutions sign with the publishers, limiting access to students and faculty only, mean less access for the public. When hard copies were available, anyone could read, digest, copy relative sections, etc. Now, depending on the institutional policy, it may be difficult even to get access to view a full text. Needless to say, the cost of individual articles is exorbitant.* [MF1]

The library would allow the public access to its online journals at a small number of "public" terminals (for those designated as "walk-ins" in the contract), but as this participant noted,

> *[The university libraries] control [public access] very, very, very tightly. I went in and asked if I could use the computer.... I had 6 articles I wanted to see, and they let me do that.* [MF1]

She concluded that this represented "the publisher's stranglehold on the universities" (MF1). Another participant had caught wind of a hopeful agreement that would greatly increase access to medical research, not only for him, but the public at large:

> *I was talking to a librarian here at the public library and she said in [British Columbia, Canada] they were...trying to link the public library with the university library so that anybody that has a library card [could view the journals].* [MM6]

It was PubMed's clear identification of, as well as links to, the open-access articles that was the one aspect which participants consistently valued. The participants felt strongly about the importance of being able to read the full text, and while a few had access to research libraries (McGill and University of British Columbia cited), others were willing to consider paying for that access. But in all cases, the value of open access to this literature was judged as a critical aspect in becoming better informed about what the research had to offer their professional practice.

Send-To Tools

Users are able to select, organize, and save the citations they find by sending the selected citations to a file, text, printer, clipboard, email, or RSS feed. Five of the participants were very impressed by how this tool allowed one to build a personal set of research resources:

> *I've done this before [email search results to self] when I have found research on the Web. I think it's great; it is very easy, and I like to keep things in my inbox. This is a great tool that I would use.* [CM2]

Another chiropractor commented on how it enabled him to integrate PubMed into his work:

> *Over the course of the day, I can sort it and then I have my list in order. Now I can take that display list and send it to text.* [CM11]

Search History

PubMed keeps a record of each search that a user conducts, under History, allowing users to not only review their previous searches but to combine them using Boolean operators. "Wow. That's amazing. If I ever want to go back to what I did, this is how. It's like you don't have to write anything down because it is all recorded automatically. This is really, really great" (HF3). While this tool only made it into the sessions of 10 participants, only one of those participants proved relatively indifferent to its powers: "That makes sense" (CM18).

Bookshelf

PubMed has integrated a number of full-text medical and life science books that can be searched separately or in conjunction with an article abstract to gain background information about key terms in the abstract (which are highlighted with links to the relevant passages in the books). Only one participant saw how access to the books might be used for providing background: "I really like it; if I am writing a thesis or a paper, I would see this as very useful" (MF5). More frequently, the participants in the study made a number of observations about the currency and focus of the research articles as the reasons they preferred them over books: "Well that's interesting [referring to Bookshelf], but I like articles because they are more current" (MF1), and "I like journal articles because they are more specific" (HF3).

A chiropractor reinforced this idea, to a degree that suggested how the background role of PubMed's Bookshelf program was being missed by some:

> *That's interesting [referring to Bookshelf]. But...once you get into reading research articles, they are much better than reading books. Books are always behind a few years.... There is a lot of bias behind the author of the book—reading an article, you get the actual truth.* [CM14]

Such responses suggest that the participants were picking up ideas about the importance of currency and the leading edge role of journals in the life sciences, ideas that need to be refined, certainly, but that are not far removed from common thinking within the research culture.

MeSH Vocabulary Database

The MeSH Database enables readers to look up the controlled vocabulary used to index the research literature. Among the 10 participants introduced to MeSH, there were challenges initially in comprehending it, to a degree not experienced with the other tools and services: "I don't understand; what is this [MeSH Database]? Is this everything with the keywords?" (CF12), and "Well, I have taken the tutorial [on the MeSH Database], and I've tried to use it, and I know it provides you with headings" (MF1). However, with further explanation and a chance to work with it, at least one participant found it "totally awesome and very useful" (MF8), and another appreciated that "it helps to distinguish things, and so you can go in one direction or another depending on what you want" (MM9). However, that same participant pointed out that what he felt was an element of bias:

> *The word "subluxation"...means something quite different to a medical doctor than it does to a chiropractor when it comes to small tiny restriction in motion or small tiny misalignment. But the medical profession still doesn't believe such things can occur, so I see here that there is no such thing here as subluxation in the sense that the chiropractor means it—very interesting.* [MM9]

PubMed's sharing of its vocabulary used for indexing also spoke to the value of this service as a source of engagement and reflection.

Email Alert and Notification Services

The National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM) is responsible for PubMed and offers users a personalized service that includes email alerts when materials of interest to them are indexed. While this tool came up in only five of the sessions, there was recognition of its value by all the participants: "I have an automatic referral for [from] NCBI every so often about updates of new research in this area" (MF1). A chiropractor had signed up with a similar service, PubCrawler, developed by a genetics research lab at Trinity College, Dublin, for alerting readers to new articles in areas of interest: "I've signed up to PubCrawler, which alerts me to new research in my area, and so [I] quite often scan these [almost daily]" (CM11).

Author Links

The ability to find other works by the same author, simply by clicking on the author's name, was not generally valued by the participants with the exception of the homeopath and one of the chiropractors: "So I have heard of him before. I would think it would be interesting; it would be interesting to know what his perspective was on other things" (CF12). With the six participants who encountered this feature divided between interest in and indifference toward it, it may seem that this way of focusing on the work of an individual researcher is not as valued as the other means of tracking work.

Comprehending Research

In the course of reflecting on PubMed tools and services, some of the participants reflected on their understanding, as well as their critical reading and use, of the research they were searching. For example, a number of participants commented on the value of reading the abstract as a necessary first step in reading research: "I wouldn't read a paper without an abstract" (HF3). By the same token, the abstract was not enough: "Definitely need to read the article. [Abstract] does not tell us enough here" (MF8). A chiropractor noted that the figures and illustrations available in the full text were needed to make sense of the orthopedic tests used in the study:

> *[The abstract is] a good summary, but...if the article had pictures and stuff, because maybe these are new tests that have come out that the school hasn't taught us yet, and they might be better than the ones we've learned...[then access] would be important.* [CM15]

Another chiropractor spoke of the importance of seeing the research methodology described in detail:

> *The abstracts usually don't show enough about the methods of an article.... It depends on the method, if the paper is actually good.* [CM14]

Similarly, a third chiropractor, who was satisfied with the abstract alone when it came to "advice or recommendations," felt that when the "treatment" of a client was at issue, then more than the abstract was needed:

> *I can't break down the methods they used [with the abstract alone], and if there's any flaws in it...I won't get it from the abstract. So I would like to see the whole thing.* [CM18]

The participants in this study were aware of their limits in understanding the research, which concerned specific domains of knowledge, as this chiropractor made clear: "I am trying to find one that is not as 'biochemistry' to read; those ones are terrible to read." Yet, within a given search, the same participant was able to find research studies which, in dealing with "exercise for treating fibromyalgia," were extremely helpful and comprehensible to the point where he "would want to prescribe" those exercises to his clients (CM16). More than one registered massage therapist was prepared to challenge the limits on her own understanding: "There's a lot of qualitative research out there on complementary therapies, but not a lot of quantitative [work], so I am looking for quantitative-oriented stuff" (MF5). However, she realized that she was not clear on what "the *p* factor" (*sic*) was and stated that "I'm not really up with statistics.... I can understand some but not all" (MF5). A second registered massage therapist was willing to push her learning beyond the concerns of clinical practice: "Well, it's not useful for me, for what I do, but it does tell me some information that I didn't know before: the cell type that they are talking about, the engagement of it" (MF7).

If participants felt somewhat overwhelmed by the research at times—"It's a little above my head" (CM13)—they also knew how to direct their reading in relation to their strengths. As a registered massage therapist put it: "This is a little too theoretical and general...[while I'm] looking for protocols for treating low back pain.... I'd be more on 'protocol,' 'orthopedic,' 'massage therapy'" (MM9). Or, as a chiropractor explained: "It's a bit too heavy; it is going to go beyond what I need clinically" (CF10). What was perhaps most notable about the participants'

responses to the research was their perception of the value of this public resource to their professional practice: “It’s pretty encouraging to find this [material in PubMed]; it’s not intimidating” (MF7), and “I use the database and feel quite comfortable with it” (CF10).

A Growing Research Culture

This research on CAM practitioners’ use of a life sciences research index reflects a growing influence of a research culture among the health professions and the public. Two participants made reference to their clients bringing in online health information:

> *My problem is telling [my clients] that what they are finding in Google is not that sound scientific information because any practitioner out there, anybody can put something out onto Google or their office website, but it’s not necessarily valid. I always have clients who are asking me about information that has come out on certain treatments, so that is primarily what I would tend to use this [PubMed] for.* [MF5]

That same increased interest in recent research had been expressed within the CAM community as well, as the administrator of a chiropractic organization stated:

> *Our membership expects us to know things instantly if they are in the public domain. They say, "Why don’t you know? Everyone else knows”.* [CM11]

The increasing prominence of research in the health services sector was also reflected in the training received by graduates of the Canadian Memorial Chiropractic College:

> *This school is very evidence-based. So a lot of the assignments are pretty much looking up research: sensitivity, specificity, tests, prognosis.... If you are efficient with any of these search engines, you can find the information quite quickly. I have used it in university as well, but not to a great extent. When I got here I had to use it much more. It depends on the clinician that you get; some are really, really, really, evidence-based, meaning they follow everything the research says.* [CM17]

Another participant noted:

> *They [at the Canadian Memorial Chiropractic College] talk so much about evidence-based care in medicine and chiropractic, and you have to have research articles in order to attempt to that. You need a way to do it and PubMed is the only way I know.* [CM16]

He went on to add that “it is free.” This was further confirmed by a classmate: “They [the school] thrust that down our throats: ‘Make sure you are up to date on that’” (CM15).

In turn, CAM practitioners are participating in the debates on the growing prominence of certain forms of medical research but in the following example also demonstrate limited understanding of the methodology:

> *I wrote an article about [how]...it is more or less impossible to do a [randomized clinical trial] of chiropractic. You can’t blind a practitioner; it’s more or less impossible to blind the patient, so any effect of randomization is nonsense.* [CM11]

Overall, there is a growing global awareness of the important role played by research in the life sciences, and these CAM practitioners are learning that this body of knowledge has much to offer, with PubMed providing one means of obtaining access to it: “There’s a lot going on in Europe and Asia that I know nothing about unless they publish in *Lymphology* or unless they present at the ISL [International Society of Lymphology]” (MF7).

Continuing Use of PubMed

In response to follow-up emails that were sent out some weeks after the session with the researcher, seven participants affirmed that they were using or were planning to use PubMed:

> *Yes, I have used PubMed since our last conversation. I use it on a weekly basis. I use it to find information for patients and for myself. I am in the process of developing an acupuncture-related website and writing a book on health-related topics.* [CM2]

Issues of access to research articles persisted in the use of PubMed for this chiropractor:

> *I have not purchased any articles, although I have been tempted. I tend to look at the full-text articles, and I have used the links LinkOut and Bookshelf with success.* [CM2]

As well, he had begun to see how he could best use PubMed in his practice:

> *I think I’ll use it to get research, to get on-the-fly research to answer questions. But I wouldn’t send my patients to use the database or conduct searches on this database after a consult with them because I think it would be too complicated. I would not refer them to PubMed to find their own research.* [CM2]

There were also indications from the participants of PubMed raising the quality of knowledge that informs their practice: “I’m using [PubMed] now to look something up instead of Google” (CF12).

Other participants spoke in the follow-up emails of their plans for continuing use of PubMed:

> *Now that I know more about it, I think I’ll use it more.... Now that we have had this session, I’ll be able to understand it a little bit better, especially this "Related Articles".* [CM15]

This same participant had explained in his session with the researcher how the Limits tool would serve his practice: “[This] would be great if I had a patient come in with a problem, and I had to really specify my search according to his needs” (CM15).

Another chiropractic student spoke of PubMed use in terms of its public access as a source of professional development:

> *I am not going to be in academia much longer. It would be great to know as much as I could before I leave. I plan on using this when I leave as well.* [CM17]

Of course, not everyone who participated in this study went on to use PubMed afterward:

> *I often learn about research studies from [online] medical news sources [eg, Medical Post] and access them through links provided in the article versus searching PubMed. There is so much information out there, and I find it hard to find time to access it all!.* [HF3]

A similar theme was sounded by a massage therapist:

> *I have no urgent reason to [use PubMed], and my time is too occupied to browse.... Currently, I do not [need PubMed].* [MM9]

And a chiropractor wrote to say that while he is sometimes "directed to PubMed as a resource" at his school, the problem is that he "could never really get the entire PDF files" that he wanted (CM13).

Discussion

This study confirmed our hypothesis that CAM practitioners would find that certain PubMed tools and services, on being introduced to them, had the potential to contribute to their engagement with and understanding of the research that interested them. This study did not seek to measure the actual contribution of these tools and services to the users' learning. It was designed to establish whether these users perceived the tools and services as having value, as a first step to subsequent studies that will assess the differences that individual tools might make to the participants' learning.

Among the PubMed tools and services to which the participants were introduced, only the service of full-text access proved a positive asset for engagement, comprehension, evaluation, and utilization. As well, the Boolean Search and Related Articles were also strong contributors to the participants' work with PubMed, and all tools and services impressed more than a few participants as to their value and contribution.

In terms of the specific design of PubMed, it is clear that the NLM is continuing to improve its design and functionality [25]. For example, the participants found MeSH to be a particularly difficult tool to grasp; however, the NLM has since provided three animated tutorials (eg, "Searching with MeSH") with voice-over, which make its operation much clearer, with at least one recent study attesting to MeSH as the preferred search strategy [26]. Still, one of this paper's reviewers wisely advised that training in MeSH would greatly improve user searches, and the ways in which a user's health vocabulary contributes to recognition and comprehension has been the subject of a recent study [27].

There have been design improvements, as well, with Related Articles, Limits, and other tools. With Bookshelf, the only feature that left over half of the participants who tried it seeing little value in it, PubMed has withdrawn the book links from an article's abstract, while still offering users the ability to search its book holdings. Yet, these medical and basic science books could well serve as background resources when a reader is stymied by concepts in biochemistry and statistics, to name two areas that this study's participants identified as weak for them [28]. There is reason, then, for the NLM to continue exploring ways of integrating the books into the literature searches in a way that encourages professionals in the health services to seek background clarification and context, as a way of improving their engagement with the research.

The study also has bearing on what has developed, with the rise of the Internet, into the "access issue." This goes back to at least 1999, when Harold Varmus, then director of the National Institutes of Health, proposed that all medical research journals make their work freely available in PubMed Central 6 months after publication, an initiative that met with the successful resistance of commercial and society publishers [29]. More recent measures include the National Institutes of Health Public Access Policy, which asks funded researchers to voluntarily deposit copies of their published work in PubMed Central, but which has managed to achieve a less than 4% compliance rate to date [30]. As well, the proposed Federal Research Public Access Act of 2006 is currently before the Senate, which would mandate open access within 6 months of publication for a large proportion of US government–funded research [31]. The results of this study would seem to support efforts to increase access to research by making it clear how greater access could contribute to the practice of CAM, as well as to more traditional forms of clinical practice.

The participants' consistent interest in having access to the full text of the articles indexed in PubMed leads to a strong and immediate recommendation to improve that access: The NLM should make every effort to capture an otherwise missing and substantial source of open access to research and scholarship, namely, the published health sciences research that has been posted by authors in institutional repositories and on websites. While PubMed has an excellent system for identifying and connecting to open-access articles made available by publishers, it needs to develop similarly effective systems for tapping the published work that authors have posted, with the publisher's permission, in archives and on websites. With numerous archiving mandates, both in place and pending, for this form of open access to research that has been funded by governments and foundations, PubMed needs to ensure that it is able to take advantage of this movement to greater openness. For example, PubMed could include a means of searching the more than 800 repositories worldwide.

For those who study the use of health information, we see a need to push beyond the obvious limitations of this study. There are sufficient grounds for further studies carefully designed to assess the contribution that particular aspects of these tools provide for different types of users, with a focus on how the design of the tools and services augments the users' comprehension, evaluation, and utilization of the materials they encounter.

The prospect of the ongoing development of resources such as PubMed, along with an increasing degree of public and

professional expectation of access to research and scholarship, holds much promise for the continuing educational quality of the Internet and society at large. While only a fraction of CAM practitioners and a proportion of their clients will pursue these forms of knowledge, there is a larger point to this. Through the combined efforts and commitment of the NLM, a good number of life science journals, and the researchers themselves (who post open-access copies of their work), the quality of knowledge that is publicly and universally available is increasing, adding to people's understanding as well as to the health care practices that affect their lives. This study is one small demonstration of how the benefits of this greater access extend to a larger community than has been commonly considered when it comes to public resources such as PubMed.

Acknowledgments

We wish to express our appreciation to Trish Dryden, Rona Achilles, and Anne Taylor-Vaisey for helping to facilitate this study. We would also thank JMIR's editor and the paper's two reviewers, as well as Anne White, who provided helpful suggestions for improving the paper, and we acknowledge the generous support of the Social Sciences and Humanities Research Council of Canada for this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Articles sent to participants prior to PubMed sessions (pdf) [PDF file (Adobe Acrobat), 50 KB - jmir_v9i2e19_app1.pdf]

Multimedia Appendix 2

Searches conducted by participants in PubMed sessions [PDF file (Adobe Acrobat), 57 KB - jmir_v9i2e19_app2.pdf]

Multimedia Appendix 3

Complete interview transcript of PubMed sessions [PDF file (Adobe Acrobat), 675 KB - jmir_v9i2e19_app3.pdf]

References

1. Lippman H. Can complementary and conventional medicine learn to get along? Bus Health 2001 Oct;19(9):15-19. [Medline: 11642053]
2. Fontanarosa PB, Lundberg GD. Alternative medicine meets science. JAMA 1998 Nov 11;280(18):1618-1619 [FREE Full text] [Medline: 9820267] [doi: 10.1001/jama.280.18.1618]
3. Vincent B, Vincent M, Ferreira CG. Making PubMed searching simple: learning to retrieve medical literature through interactive problem solving. Oncologist 2006 Mar;11(3):243-251 [FREE Full text] [Medline: 16549808] [doi: 10.1634/theoncologist.11-3-243]
4. Song YL, Yang XZ, Lu XY. Discussion on methods for searching traditional Chinese medical literature in English in PubMed-MEDLINE database. Zhongguo Zhong Xi Yi Jie He Za Zhi 2006 Jul;26(7):651-654. [Medline: 16983926]
5. ; Royal College of Nursing. Report of Key Findings of RCN's Survey of the Information Needs of Nurses, Health Care Assistants, Midwives and Health Visitors. London, UK: Royal College of Nursing; 2005. URL: http://www.rcn.org.uk/downloads/news/INA%20report%20external.doc [accessed 2007 Jun 21] [WebCite Cache ID /5PutqHFAm]
6. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [Medline: 16867972] [doi: 10.2196/jmir.8.2.e9]
7. Friedman DB, Hoffman-Goetz L. A systematic review of readability and comprehension instruments used for print and web-based cancer information. Health Educ Behav 2006 Jun;33(3):352-373. [Medline: 16699125] [doi: 10.1177/1090198105277329]
8. Carolan M. Health literacy and the information needs and dilemmas of first-time mothers over 35 years. J Clin Nurs 2007 Jun;16(6):1162-1172. [Medline: 17518891] [doi: 10.1111/j.1365-2702.2007.01600.x]
9. 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 Apr 21;8(2):e6 [FREE Full text] [Medline: 16867969] [doi: 10.2196/jmir.8.2.e6]
10. Tannery NH, Wessel CB, Epstein BA, Gadd CS. Hospital nurses' use of knowledge-based information resources. Nurs Outlook 2007 Jan;55(1):15-19. [Medline: 17289463] [doi: 10.1016/j.outlook.2006.04.006]
11. Dee C, Stanley EE. Information-seeking behavior of nursing students and clinical nurses: implications for health sciences librarians. J Med Libr Assoc 2005 Apr;93(2):213-222 [FREE Full text] [Medline: 15858624]

12. Bennett NL, Casebeer LL, Zheng S, Kristofco R. Information-seeking behaviors and reflective practice. J Contin Educ Health Prof 2006 Dec;26(2):120-127. [Medline: 16802313] [doi: 10.1002/chp.60]
13. Sood A, Erwin PJ, Ebbert JO. Using advanced search tools on PubMed for citation retrieval. Mayo Clin Proc 2004 Oct;79(10):1295-9; quiz 1300. [Medline: 15473412]
14. Sood A, Ghosh AK. Literature search using PubMed: an essential tool for practicing evidence- based medicine. J Assoc Physicians India 2006 Apr;54:303-308. [Medline: 16944614]
15. Eldredge JD. Search strategies for population and social subjects in a medical school curriculum. Med Ref Serv Q 2004;23(4):35-47. [Medline: 15778181] [doi: 10.1300/J115v23n04_04]
16. Saxton JD, Owen DJ. Developing optimal search strategies for finding information on herbs and other medicinal plants in MEDLINE. J Altern Complement Med 2005 Aug;11(4):725-731. [Medline: 16131300] [doi: 10.1089/acm.2005.11.725]
17. Bardia A, Wahner-Roedler DL, Erwin PL, Sood A. Search strategies for retrieving complementary and alternative medicine clinical trials in oncology. Integr Cancer Ther 2006 Sep;5(3):202-205. [Medline: 16880424] [doi: 10.1177/1534735406292146]
18. 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]
19. Johnson ED. Some benchmark searches for testing search capabilities and medical coverage of internet discovery tools. J Med Internet Res 2000 Sep 27;2(3):E19 [FREE Full text] [Medline: 11720942] [doi: 10.2196/jmir.2.3.e19]
20. Alexander PA, Jetton TL, Kulikowich JM. Interrelationship of knowledge, interest, and recall: assessing a model of domain learning. J Educ Psychol 1995 Dec;87(4):559-575. [doi: 10.1037/0022-0663.87.4.559]
21. Alexander PA. The past, present and future of knowledge research: A reexamination of the role of knowledge in learning and instruction. Educ Psychol 1996;31(3):89.
22. Ward LG, Kendrach MG, Price SO. Accuracy of abstracts for original research articles in pharmacy journals. Ann Pharmacother 2004;38(7-8):1173-1177. [Medline: 15150375] [doi: 10.1345/aph.1D416]
23. Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. JAMA 1999;281(12):1110-1111 [FREE Full text] [Medline: 10188662] [doi: 10.1001/jama.281.12.1110]
24. Read MEDLINE abstracts with a pinch of salt. Lancet 2006 Oct 21;368(9545):1394. [Medline: 17055921] [doi: 10.1016/S0140-6736(06)69578-0]
25. Muin M, Fontelo P. Technical development of PubMed interact: an improved interface for MEDLINE/PubMed searches. BMC Med Inform Decis Mak 2006 Nov 3;6(1):36 [FREE Full text] [Medline: 17083729] [doi: 10.1186/1472-6947-6-36]
26. Chang AA, Heskett KM, Davidson TM. Searching the literature using medical subject headings versus text word with PubMed. Laryngoscope 2006 Feb;116(2):336-340. [Medline: 16467730] [doi: 10.1097/01.mlg.0000195371.72887.a2]
27. 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]
28. Corsi AK. Books for free? How can this be? - A PubMed resource you may be overlooking. Biol Cell 2006 Jul;98(7):439-443. [Medline: 16776646] [doi: 10.1042/BC20050093]
29. Varmus H, Lipman D, Brown P. PubMed Central: An NIH-Operated Site for Electronic Distribution of Life Sciences Research Reports. 1999 Aug 30. URL: http://www.nih.gov/about/director/pubmedcentral/pubmedcentral.htm [accessed 2007 Jun 21] [WebCite Cache ID 5MgcI6eJs]
30. Stebbins M, Davis E, Royland L, White G. Public access failure at PubMed. Science 2006 Jul 7;313(5783):43. [Medline: 16825550] [doi: 10.1126/science.313.5783.43b]
31. Federal Research Public Access Act of 2006 (FPAA) (introduced in Senate). Washington, DC: Library of Congress. URL: http://thomas.loc.gov/cgi-bin/query/z?c109:S.2695: [accessed 2007 Jun 21] [WebCite Cache ID 5Mi6xd6Ws]

Abbreviations

CAM: complementary and alternative medicine
MeSH: medical subject heading
NLM: National Library of Medicine

submitted 19.02.07; peer-reviewed by R Cullen, L Lafrado; comments to author 15.05.07; revised version received 07.06.07; accepted 07.06.07; published 29.06.07.

Please cite as:
Willinsky J, Quint-Rapoport M
How Complementary and Alternative Medicine Practitioners Use PubMed
J Med Internet Res 2007;9(2):e19
URL: http://www.jmir.org/2007/2/e19/
doi:10.2196/jmir.9.2.e19
PMID:17613489

Original Paper

Using Internet and Mobile Phone Technology to Deliver an Automated Physical Activity Program: Randomized Controlled Trial

Robert Hurling[1], PhD; Michael Catt[1], BSc; Marco De Boni[1], PhD; Bruce William Fairley[2], PhD; Tina Hurst[1], PhD; Peter Murray[1], CStat, MPhil; Alannah Richardson[1], PhD; Jaspreet Singh Sodhi[2], PhD

[1]Unilever Corporate Research, Colworth, Bedford, UK

[2]Tessella Support Services PLC, Abingdon, Oxfordshire, UK

Corresponding Author:
Robert Hurling, PhD
Unilever Corporate Research
Colworth, Bedford
MK44 1LQ
United Kingdom
Tel: +44 1234 22 2992
Fax: +44 1234 24 8010
Email: bob.hurling@unilever.com

Abstract

Background: The Internet has potential as a medium for health behavior change programs, but no controlled studies have yet evaluated the impact of a fully automated physical activity intervention over several months with real-time objective feedback from a monitor.

Objective: The aim was to evaluate the impact of a physical activity program based on the Internet and mobile phone technology provided to individuals for 9 weeks.

Methods: A single-center, randomized, stratified controlled trial was conducted from September to December 2005 in Bedfordshire, United Kingdom, with 77 healthy adults whose mean age was 40.4 years (SD = 7.6) and mean body mass index was 26.3 (SD = 3.4). Participants were randomized to a test group that had access to an Internet and mobile phone–based physical activity program (n = 47) or to a control group (n = 30) that received no support. The test group received tailored solutions for perceived barriers, a schedule to plan weekly exercise sessions with mobile phone and email reminders, a message board to share their experiences with others, and feedback on their level of physical activity. Both groups were issued a wrist-worn accelerometer to monitor their level of physical activity; only the test group received real-time feedback via the Internet. The main outcome measures were accelerometer data and self-report of physical activity.

Results: At the end of the study period, the test group reported a significantly greater increase over baseline than did the control group for perceived control ($P < .001$) and intention/expectation to exercise ($P < .001$). Intent-to-treat analyses of both the accelerometer data ($P = .02$) and leisure time self-report data ($P = .03$) found a higher level of moderate physical activity in the test group. The average increase (over the control group) in accelerometer-measured moderate physical activity was 2 h 18 min per week. The test group also lost more percent body fat than the control group (test group: −2.18, SD = 0.59; control group: −0.17, SD = 0.81; $P = .04$).

Conclusions: A fully automated Internet and mobile phone–based motivation and action support system can significantly increase and maintain the level of physical activity in healthy adults.

(J Med Internet Res 2007;9(2):e7) doi:10.2196/jmir.9.2.e7

KEYWORDS

Behavior change; Health behavior; Behavior therapy; Obesity prevention; Health promotion; Exercise; Cellular phone; Internet; Consumer health informatics; Randomized controlled trial

Introduction

Physical inactivity is a major concern for developed societies. It accounts for about 12% of all deaths [1] and is associated with debilitating conditions [2]. In contrast, physical activity has been linked to positive health outcomes [3] and general well-being [4].

Government organizations recognize physical activity, along with a healthy diet, as playing an important role in the prevention of obesity [5], with a recommended level of moderate physical activity for adults of at least 30 min on most days of the week [6,7], with high-risk individuals benefiting from tailored interventions [8]. Unfortunately, infrequent exercise participation is common [9], starting even within late adolescence [10].

Internet-based behavioral change interventions minimize face-to-face interaction, thereby increasing cost-effectiveness through greater accessibility [11]. Partially automated programs, in which advice from therapists is delivered via email, can help people change health behaviors [12,13], and fully automated telephone counseling systems have also increased self-reported physical activity [14]. However, no research has evaluated the effect of a fully automated Internet-based system, with real-time objective feedback, on physical activity over several months [15]. Longer term studies using pedometers often rely on self-report [16].

Internet-based physical activity websites differ in their level of interaction, from individually tailored assistance to general guidelines or advice [17]. While more interactive elements (such as emailing weekly lessons) improve the number of people achieving health-related behavior change goals [13], it is still the case that Internet- and email-based systems can fail to hold participant interest [18]. A comparison of similar systems with different levels of interactivity found that the more interactive system was better able to retain participants [19].

We have developed a fully automated Internet, email, and mobile phone system [19] based on a range of social psychological theories (Social Comparison [20], Decisional Balance [21], Elaboration Likelihood [22], and Goal [23]). We used a Bluetooth [24] connected wrist-worn accelerometer to measure physical activity and provide feedback to participants.

Our primary hypothesis was that a group provided with access to the Internet and a mobile phone–based physical activity program would maintain a higher level of physical activity over 9 weeks than a control group who wore physical activity monitors but received no feedback and had no access.

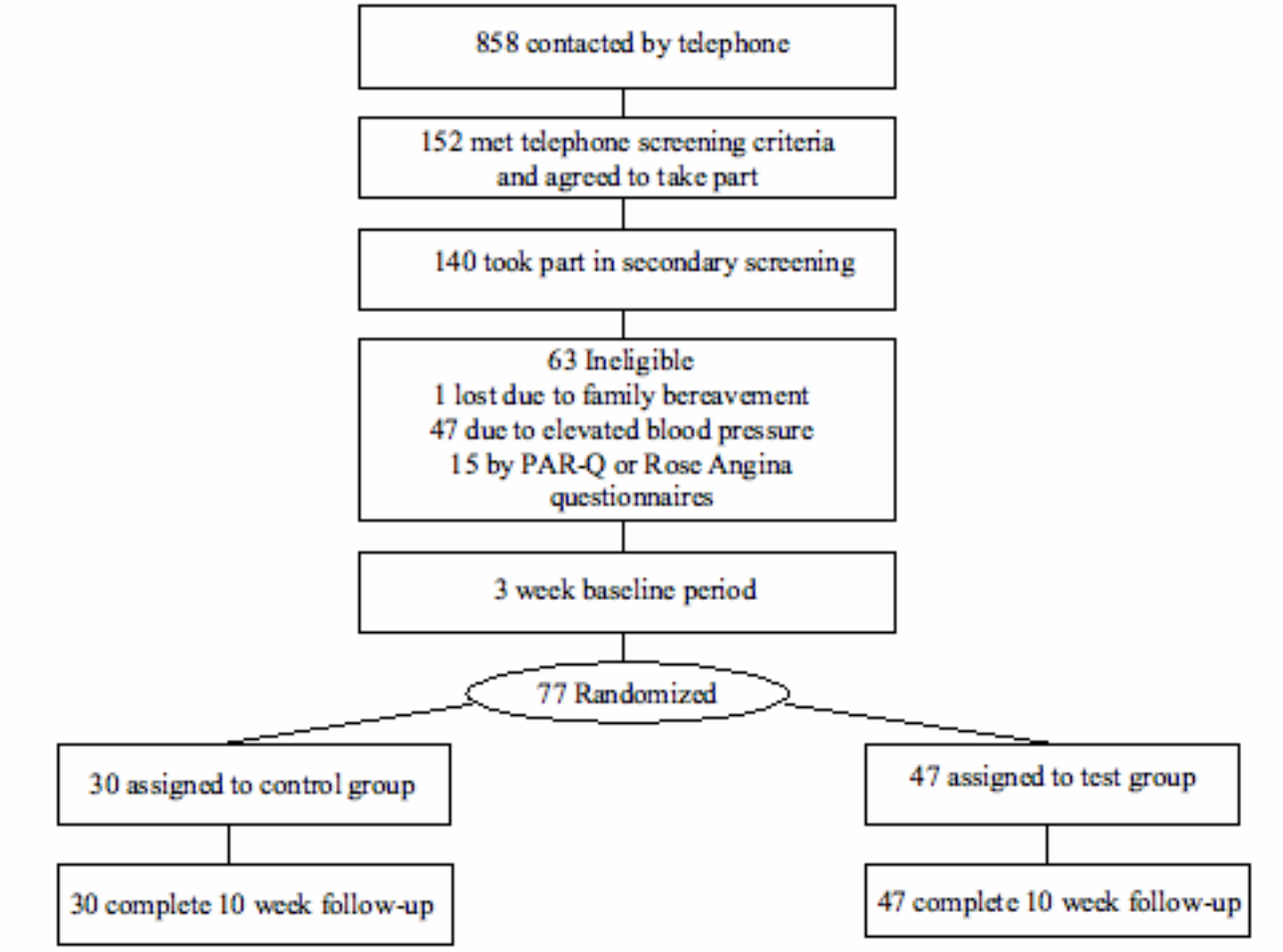
Methods

Participants

A total of 140 people were initially recruited via a market research recruitment agency and passed the telephone screening (Figure 1) with self-report eligibility criteria as follows: age 30 to 55 years; body mass index (BMI) 19 to 30 (calculated from reported height and weight); not vigorously active; not taking regular prescription medication; Internet and email access; mobile phone user; and not employed by Unilever.

All participants agreed not to take part in any other studies, were briefed on the study aims, and signed an informed consent form in accordance with the Declaration of Helsinki [25]. Participants scoring one or more items on the Physical Activity Readiness Questionnaire (PAR-Q) [26] or the Rose Angina Questionnaire [27] were not accepted into the study and were provided with a letter to seek medical advice from their general practice doctor. For example, exclusion criteria from these questionnaires included the following: a heart condition, pain in the chest when exercising, and a joint problem that might be aggravated by exercise. In total, 77 healthy adults (51 women, 26 men) between 30 and 55 years (mean = 40.4 years; SD = 7.6) with a mean BMI of 26.3 (SD = 3.4) took part in the study, all living within 50 km of the study center in Sharnbrook, Bedfordshire, United Kingdom.

Figure 1. Flow diagram of study participation



Design

The 77 participants came to the center and were issued a wrist-worn accelerometer and Bluetooth-compatible mobile phone (Nokia 6230, with their SIM card inserted). After 3 weeks of monitoring baseline physical activity, participants returned and were stratified by age, gender, and BMI and were randomly allocated to either the control (n = 30) or test group (n = 47). More participants were allocated to the test group in order to maximize information on use of the system. All participants received £30 for attending the initial screening at the center, £140 to cover mobile phone costs, and £290 at closeout.

Physical Activity Monitor

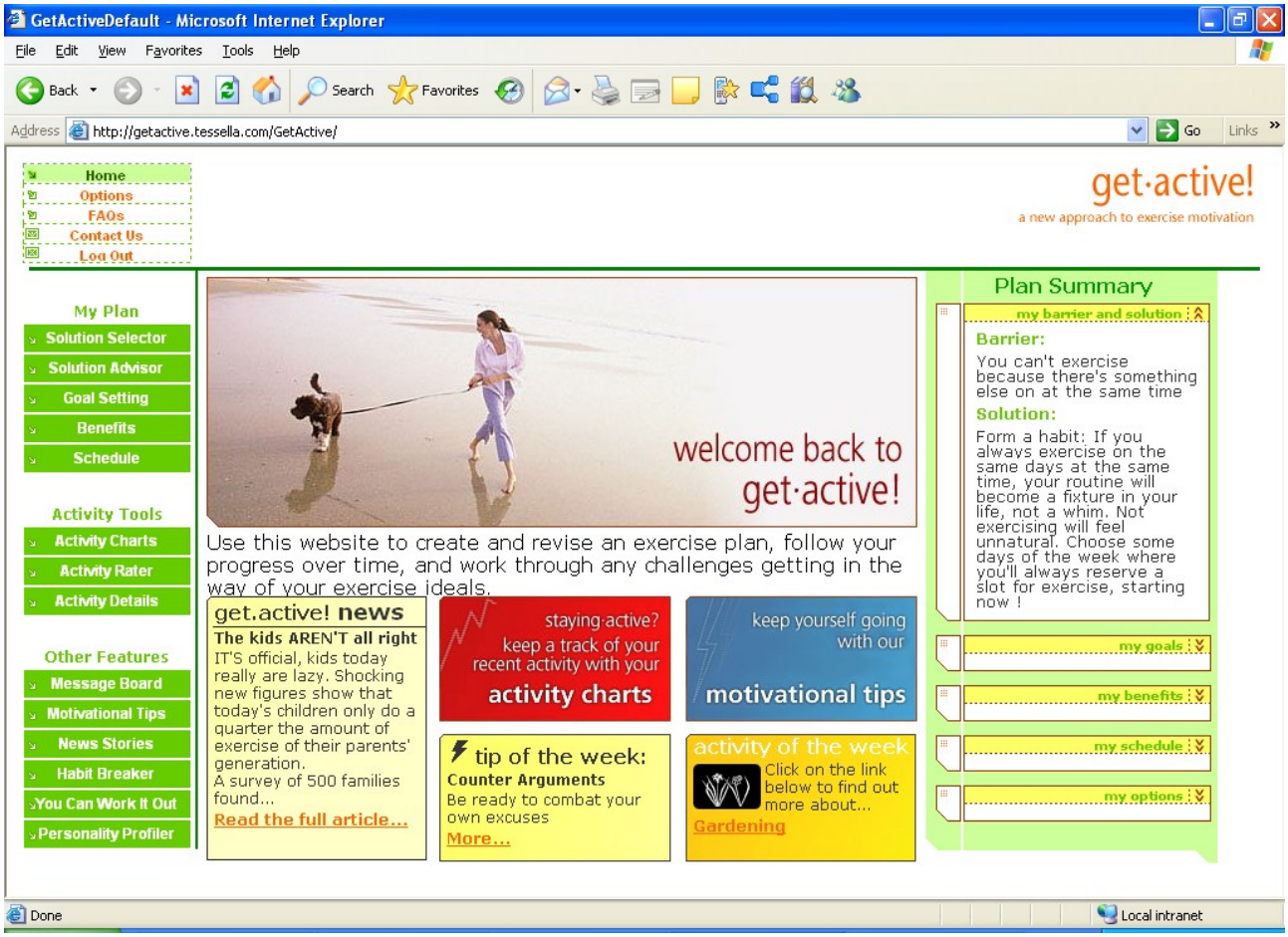
Although pedometers are low cost, they are typically attached to a waist band and therefore primarily record walking, making 24-hour monitoring more difficult. In contrast, accelerometry tools record a wider range of movement and have greater flexibility for body positioning, allowing for sustained monitoring even during sleep. Accelerometers have been widely used to monitor physical activity [28-30], both for school-age children [31] and adults [32].

Together with a technology company [33], we developed a Bluetooth wrist-worn device (Bluetooth Actiwatch) containing a miniature uniaxial accelerometer unit recording all movement over 0.05 g, excluding readings outside 3 to 11 Hz to eliminate gravitational artifacts. The signal was measured 32 times per second and digitally processed to integrate both the amount and duration of movement. Data were recorded with an epoch resolution of 2 min. The typical battery life was 6 months.

Behavior Change System

The Internet, email, and mobile phone behavior change system (Figure 2) was similar to that used in previous studies [19]. An introductory series of screens (Multimedia Appendix) helped test participants identify their perceived barriers to physical activity, offered solutions, and advised on appropriate wording for a commitment [34]. A weekly series of screens (see the Multimedia Appendix) asked the participants to report their exercise level during the last week, before providing constructive feedback on performance relative to their own target and the test group. The system included a weekly schedule (or diary) for planning physical activity sessions over the next 7 days (see the Multimedia Appendix), for which participants could choose to receive email and/or mobile phone reminders, an approach that has been effective in combination with implementation intentions [35]. The schedule included an automated “assessor” that provided feedback on the amount and type of physical activity being planned, advising a reduction in the case of participants planning to make very large increases compared with previous weeks.

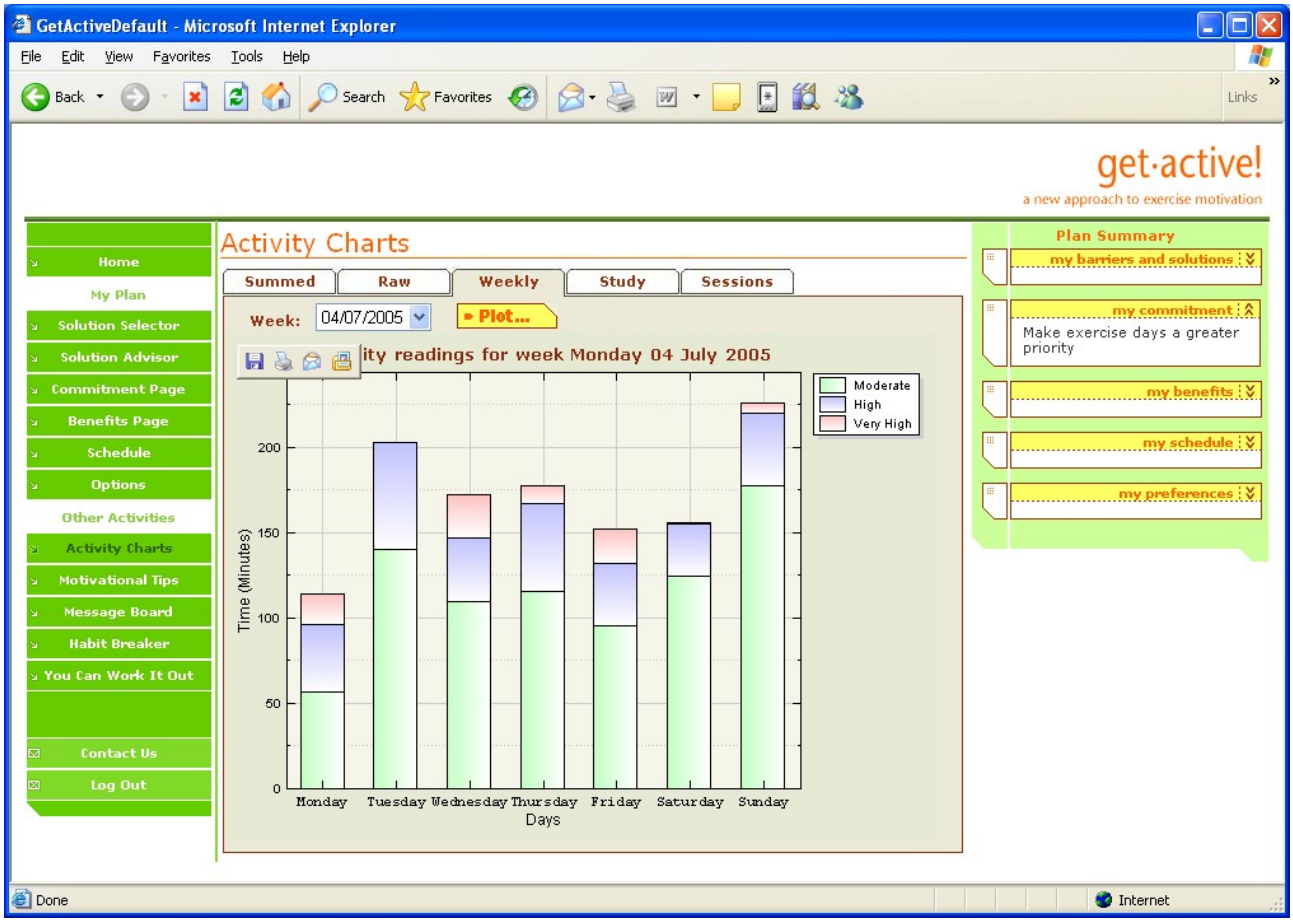
Figure 2. Behaviour change system home page



A text-based automated dialogue module helped participants identify their perceived barriers and offered tailored solutions (see the Multimedia Appendix). For example, for the barrier "You can't exercise because there's something else on at the same time," one of the solutions offered could be "Form a habit: If you always exercise on the same days at the same time, your routine will become a fixture in your life, not a whim. Not exercising will feel unnatural. Choose some days of the week where you'll always reserve a slot for exercise, starting now!" Solutions were tailored to the individual via an underlying matrix that contained a strength of association between each barrier and solution. The strength of association between solutions and barriers increased in line with the increase in the level of physical activity of participants who had previously selected them.

Participants were also encouraged to select three motivating benefits, for which email and/or mobile phone text messages were optional. There was a library of information on a range of different physical activities, from household duties to team sports, and a chat-room style message board. Charts displayed real-time output from their physical activity monitor in three bands, moderate, high, and very high, with summaries for that day, the week (Figure 3), and the total study period, including the test group average. Low-level activity was excluded from the charts to reduce background noise. The system provided motivational tips matched to each participant's current physical activity level. An automated dialogue therapy module helped people transform their rigid beliefs about exercise into more flexible, helpful beliefs [34,36]. The automated dialogue therapy module guided the participants through a series of steps, identifying a situation when a planned exercise was not carried out, identifying a "reason" the exercise was not carried out, explaining that this reason is actually a belief, and describing the difference between flexible and rigid (inflexible) beliefs, helping the participant create a more flexible belief to use next time he or she is in the same situation.

Figure 3. Weekly activity charts



Procedures

At the study center, participants received a full explanation about all procedures and were given an opportunity to ask questions. They were instructed to wear the Bluetooth Actiwatch on the wrist of their nondominant arm continuously for the following 12 weeks (3 weeks baseline and 9 weeks intervention). As the accelerometer was not fully waterproof, participants were asked to remove it when washing, bathing, or swimming. Following collection of 3 weeks of baseline data, the test group participants received a short demonstration of the Internet-based behavior change system; the control group also came to the center but only received verbal advice on recommended physical activity levels. The test group then had access to the Internet-based behavior change system for 9 weeks, whereas the control group had no access and received no feedback.

Dependent Measures

The primary dependent measure was change in moderate physical activity recorded by the longer version of the International Physical Activity Questionnaire (IPAQ) [37] and the wrist-worn accelerometer. Changes in weight, percent body fat (as measured by bioelectrical impedance scales [38]), height [39], and resting blood pressure [40,41] were secondary measures. All measures were taken before and after the 9-week intervention period in September and December 2005, respectively.

A set of cognitive items was developed specifically for the study, each scored against a 7-point numbered scale ranging from “Strongly Disagree” to “Strongly Agree.” The items to measure motivation were as follows: “I am very satisfied with my level of fitness,” “I am very satisfied with my current level of motivation to exercise,” “I consider myself to be very healthy,” and “I am very happy about my general level of well-being” (Cronbach alpha = .89). The items to measure perceived control were “Exercise is too much effort” and “I feel in control of how much exercise I get” (Cronbach alpha = .63), and those measuring intention/expectation to exercise were “I intend to exercise for 30 minutes at least 3 times in the next week” and “I realistically expect that I will actually exercise for 30 minutes at least 3 times in the next week” (Cronbach alpha = .92). One item measured participant interest in using an Internet-based behavior change system: “I think an Internet-based motivation program could help people to take more exercise.”

Participants also completed an exercise Skills and Knowledge Questionnaire that asked about skills used to increase physical activity [42]. Factor analysis indicated that participant responses fell into those related to their external environment (an external factor) and those related to internal motivation and confidence (an internal factor) [43]. The external factor items were “How confident are you that you know how to use prompts (reminders) to increase your physical activity,” “How confident are you that you know how to use rewards to increase your physical activity,” “How confident are you that you know how to get support from your family to increase your physical activity,” and “How confident are you that you know how to get support from your

friends or colleagues to increase your physical activity" (Cronbach alpha = .89). The internal factor items were "How confident are you that you know how to set yourself achievable goals to increase your physical activity" and "How confident are you that you know how to make an action plan to increase your physical activity" (Cronbach alpha = .91).

Statistical Analyses

Three participants were found to have faulty Actiwatches and so were removed from all statistical analyses. IPAQ data were processed according to the Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire [44]. We used an analysis of covariance model for the posttreatment data with pretreatment as a baseline covariate. A square-root transformation was applied to the data to protect against non-normality.

Actiwatch data for contiguous periods of zero extending longer than 30 min were omitted, as were data for periods with > 5000 counts for longer than 10 min since both these conditions indicated temporary malfunctioning of the accelerometer. Only days that had at least 10 h of recorded data following these corrections were retained for analysis [45]; of the 6634 eligible days of data, 177 were dropped (less than 3%).

A Generalized Estimating Equation Model with log link and Poisson distribution was used to calculate the number of 2-min epochs spent within three metabolic equivalent (MET) ranges [46], corresponding to moderate intensity (MET level over 3 and up to 6), high intensity (MET level over 6 and up to 9), and very high intensity (MET level over 9) (personal communication, S Brage, MRC Epidemiology Unit, Strangeways Research Laboratories, Cambridge, UK, 2005), during each individual's waking day (identified from the 24-h Actiwatch data). Data points were only counted if they were part of a continuous bout of exercise of at least 10 min within the MET range. We corrected for baseline activity levels and week of study, but not for the length of day, as we were encouraging people to be more active, irrespective of the length of time they were awake. In order to represent the underlying signal, the data were smoothed using a moving average filter of width ± 1 point. Modifying the width of the filter had little effect on the results of the analysis.

We focused on the difference in total time of nonsedentary physical activity between the two groups rather than the absolute amount of physical activity within each group, as estimates of the latter can vary by a factor of 10 depending on the threshold point used [47].

Participants were instructed to remove the accelerometer for swimming—an activity selected by 36% of the test group who logged on. Therefore, our accelerometer-based estimate of physical activity did not fully account for all exercise undertaken, potentially attenuating any differences observed between the test and control groups. Anthropometric measures at the end of the study were analyzed using normal analysis of covariance models with baseline prestudy values as covariate. All analysis was carried out using SAS, version 9.1.3 [48].

Results

A preliminary analysis showed that there were no differences between groups for baseline measures of age, weight, BMI, percent body fat, blood pressure, or initial level of physical activity, whether measured by the Actiwatch or IPAQ (Table 1).

Table 1. Baseline characteristics of participants

Variable	Test Group* (n = 47)	Control Group* (n = 30)	*P* value†
Women (%)	64	70	.63
White ethnicity (%)	100	97	.39
Household broadband access (%)	29	22	.43
Age (years)	40.5 (7.1)	40.1 (7.7)	.97
Weight (kg)	75.1 (11.7)	73.9 (10.2)	.60
Height (cm)	166.3 (6.6)	165.2 (7.7)	.38
BMI	26.2 (2.8)	26.5 (4.1)	.68
Percent body fat (%)	30.2 (6.5)	31.0 (10.1)	.52
Blood pressure (mmHg) Systolic	119.8 (7.7)	118.2 (8.4)	.40
Blood pressure (mmHg) Diastolic	78.3 (5.7)	77.9 (6.1)	.82
Actiwatch accelerometer-measured time (epochs) spent above 3 and up to 6 METs during 3-week initiation period	228.0 (52.1)	214 .2 (53.1)	.11
Initial IPAQ self-report level of physical activity (MET mins)	4350 (3200)	3868 (2257)	.44

*Values are expressed as mean (SD) except for the first three variables.

†*P* value is the probability that the two groups differ.

Website Usage (Test Group Only)

More than 85% (mean = 86.4%, SD = 2.1) of test participants logged on each week during the first 4 weeks, decreasing to a plateau around 75% (mean = 76.1%, SD = 5.1) for the last 5 weeks. This level is lower than for partially automated behavior change systems [12] but higher than for other minimal-contact interventions [49,50].

The average number of log-ons per week was 2.9 (SD = 0.5), with short average duration of 7.5 min (SD = 0.9). The most frequently used components of the system were the activity charts (showing the accelerometer feedback data), the schedule (weekly exercise planner), and chat-room style message board. All components of the system were accessed by at least 33% of the participants during the intervention period. Typically, participants quickly formed an idiosyncratic preference for a few components of the system and then repeatedly used these throughout the intervention.

Comments on the message board indicated that participants found the system both educational and motivational, for example, “I am amazed looking at the graphs sometimes — I took my little fella to Bezerks in Northampton on Thursday morning and the graph went crazy with all the running around I did!”

The most popular (frequent) benefits of exercise were “Exercise will help me with weight loss” (n = 19), “I will have more energy” (n = 13), and “I will improve my muscle tone” (n = 11). The most commonly selected barriers to physical activity were time conflicts (n = 27), low motivation (n = 11), and procrastinating (n = 6).

Self-Reported Changes in Physical Activity

As shown in Table 2, an intent-to-treat analysis of (the square-root transformed) MET minutes per week found no significant difference, after adjusting for the baseline covariate, between the test group (mean = 12.0, SE = 3.1) and the control (mean = 4.0, SE = 4.1), with $P = .12$ (95% CI for the difference = −2.3-18.3). When considering only MET minutes per week within leisure time, the increase in the test group was significantly higher than the control. The reduction in weekly hours spent sitting in the test group was significantly different from the control. There was a similar trend when breaking the data down into weekday sitting and weekend sitting (Table 2).

Table 2. Self-reported physical activity in test and control groups

Self-Reported Physical Activity Variable	Test Group* (n = 47)	Control Group* (n = 30)	*P* value†	95% CI for Difference
MET min/week				
Overall	12.0 (3.1)	4.0 (4.1)	.12	−2.3 to 18.3
Leisure time	4.1 (2.6)	−5.5 (3.5)	.03	0.8 to 18.3
Change in weekly hours spent sitting				
Overall	−5.9 (2.0)	1.4 (2.7)	.03	−14.0 to −0.5
Weekday	−5.2 (1.7)	−0.2 (2.3)	.08	−10.8 to 0.7
Weekend	−0.9 (0.6)	1.2 (0.8)	.04	−4.2 to −0.1

*Values expressed as mean (SD).

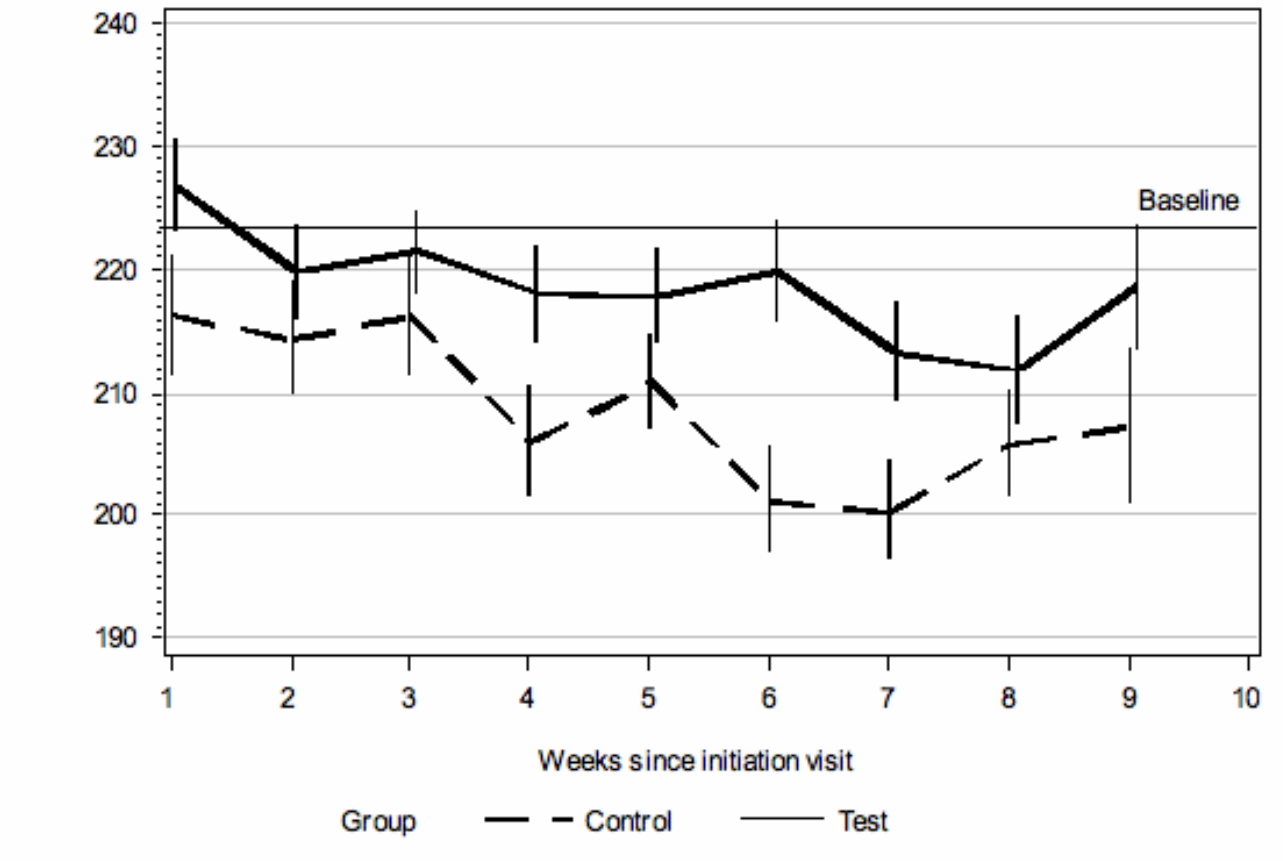
†*P* value is the probability that the two groups differ.

Accelerometer Data

We collected 4811124 data points from study participants, which represented 94.1% of the total expected based on the first and last recorded points for each individual (5114431).

Average sleep times for the two groups were not significantly different (test group: 467 min, SD = 40; control group: 468 min, SD = 38; $P = .92$). There was a significant trend over the whole study period indicating more time spent in the 3 to 6 MET range (moderate physical activity, eg, brisk walking) for the test group (log scale mean = 5.39, SE = 0.01) versus the control group (log scale mean = 5.34, SE = 0.01) with $P = .02$ (95% CI for the difference = 0.01-0.08). In the original units, this represents 218.5 epochs for the test group and 208.7 epochs for the control group, a difference of 19.7 min/day on average. The test group was consistently higher than the control group (in the moderate activity range) across all weeks of the study (Figure 4). Hence, the accelerometer data corroborated the significantly greater increase in physical activity self-reported by the test group. There was no significant difference between the groups in the ranges above 6 METs (log scale mean test group = 3.87, SE = 0.05; log scale mean control = 3.86, SE = 0.06; $P = .94$).

Figure 4. Accelerometer-measured mean number of 2-min epochs spent in moderate intensity MET range (above 3 and up to 6); error bars represent SE; baseline is 3-week average before start of intervention



Cognitive Measures

At the end of the study period, the test group reported a significantly greater increase over baseline than did the control group for the perceived control factor (mean change test group = +0.57; mean change control = −0.37; $P < .001$) and intention/expectation to exercise factor (mean change test group = +0.45; mean change control = −0.01; $P < .001$), but there was no significant difference in the motivational change factor (mean change test group = +0.11; mean change control = −0.02; $P = .56$). However, there were differences at the level of individual items within the factor. The test group rated themselves as more satisfied with their fitness (test mean = 3.94; control mean = 3.25; $P = .047$; final ratings corrected for baseline) and well-being (test mean = 5.04; control mean = 4.01; $P = .007$), a noteworthy result as changes in self-related measures of health have been related to objective health outcomes [51].

The Skills and Knowledge Questionnaire indicated that, after adjusting for baseline, the test group had a significantly higher sense of internal control (test mean = 7.24; control mean = 5.85; $P = .003$) and external control (test mean = 6.38; control mean = 5.33; $P = .01$) over exercise than did the control group at the end of the study period.

After the study period, the test group had a significantly greater interest in using an Internet-based behavior change system than the control group (test group = 4.92; control = 3.85; $P < .001$), indicating that test group participants had a positive experience.

Anthropometry

The difference between the change in the test group's BMI (mean change = −0.24, SE = 0.11) over the study period and that for the control group (mean change = 0.10, SE = 0.14) approached significance ($P = .06$; 95% CI = −0.02-0.70). The difference between the change in the test group's percent body fat (mean change = −2.18%, SE = 0.59) over the study period and that for the control group (mean change = −0.17%, SE = 0.81) was statistically significant ($P = .04$; 95% CI = 0.06-3.94). There was no significant change in either diastolic (mean change test = 0.69, SE = 1.14; control = 0.73, SE = 1.49; $P = .98$) or systolic (mean change test = 0.13, SE = 1.33; control = 0.41, SE = 1.71; $P = .90$) blood pressure.

Discussion

Increasing physical activity in the general population has an important role in the prevention of obesity and associated health problems [15,52]. We have shown that physical activity can be increased via a fully automated Internet-based behavior change system. The capture of real-time accelerometer data over 9 weeks while participants went about their everyday lives is in itself advancement for the field. Due to the use of Bluetooth technology, we were able to promptly detect and resolve most malfunctions in the accelerometers, resulting in extremely low data loss (< 10%). Access to the system was encouraging, with more than 70% of participants continuing to log on at least twice a week for all 9 weeks of the study. We attribute this to the interactive nature of the system we have developed [19]. It is

essential that automated systems engage people in order for the behavior change program to have an impact [53]. As we only compared our system with a control group who received verbal advice, we cannot conclude that it would be superior to other interventions [54].

Although we observed an increase in accelerometer-measured physical activity for the test group over the control group, our analysis was limited by its uniaxial nature; future studies could employ a triaxial accelerometer so that greater differentiation of physical activity types can be achieved [55]. Also, since our accelerometers were only splash-proof, we were not able to capture physical activity for water sports.

The difference between the test and control group accelerometer-measured physical activity was apparent for most of the 9-week intervention (see Figure 4). The control group began at the same level as the test group but then decreased to a greater extent. It is likely that both test and control groups had initially higher levels of physical activity than their norm, due to awareness of being monitored and/or completing the questionnaires [56]. This suggests that the Internet-based behavior change system enabled the test group to maintain their elevated level of physical activity. The size of the difference in physical activity between the two groups is considerable; an increase of 2 h 18 min per week represents 92% of the recommended [6] 2 h 30 min, although further work is required to clarify how absolute continuous accelerometry measurements relate to the 30 min/day government recommendation.

In line with the Theory of Planned Behavior [57], which has been widely applied to a range of health behaviors [58], we found that the more physically active test group also reported a greater perceived control over their exercise behavior and greater intention/expectation to exercise. The test group level of motivation was not significantly different from the control, indicating that the intervention primarily influenced volitional aspects of behavior [23]. Unmotivated groups may require additional modules encouraging them to engage in the target behavior [59,60].

It was clear from the verbatim comments posted by participants on the message board that the accelerometer-based activity charts acted as educational information, allowing them to link periods of high physical activity to events in their everyday life. Indeed, everyday physical activities such as walking are considered to have the greatest potential for increasing overall activity levels of a sedentary population [61], and greater awareness of personal activity levels may lead to more positive intentions [62,63]. Unfortunately, a deficiency of our study is that we did not collect sufficient qualitative data to make a thorough analysis of how participants perceived the system.

In line with other research [64,65], we found BMI a less sensitive measure for physical activity interventions. However, the test group lost significantly more percent body fat in comparison to the control group, indicating that the increased physical activity may have led to greater muscle mass. An Internet-based behavior change system that improved diet as well as increasing physical activity may lead to more substantial losses of body fat and reduced BMI [66].

Most participants selected from a relatively small set of barriers and were motivated by similar benefits, as has been reported by other researchers [67]. Further work is required to explore how other groups of potential users react to the system and whether greater personalization is required [68].

Based on a range of behavior change principles taken from the literature [20,23,69], we included many different processes within our system, and so it is hard to determine which where the most effective in helping participants to change their behavior. The most popular parts of the system were the activity charts (showing the accelerometer feedback), schedule (weekly exercise planner), and chat-room style message board. The activity charts provided participants with feedback on their performance, which may have increased awareness and acted as a motivational spur for change, in line with Goal Theory [23]. The schedule can be considered a tool for making highly specific implementation intentions, which other research has shown to be an effective intervention for behavior change [59]. The message board could be considered a modern day representation of subjective norm (social pressure), as described within the Theory of Planned Behavior [57]. However, popularity (in terms of frequency of use) does not necessarily imply greater efficacy for behavior change. It is also notable that all parts of the system were used by at least one third of participants; it may be the case that each individual requires an idiosyncratic selection of support tools to achieve behavior change such that no one tool can be universally considered the most influential. Further work is required to determine how parts of the system interact to impact individual behavior change and how to optimize the exposure period; 9 weeks may not be necessary [70].

In summary, we found that participants with access to a fully automated behavior change system engaged in, on average, 2 h 18 min more physical activity per week than those with no access.

Acknowledgments

Jaspreet Singh Sodhi had full access to all of the data and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix

Screenshots of the behaviour change system (ppt) [PPT file (MS Powerpoint), 980 KB - jmir_v9i2e7_app1.ppt]

References

1. Murray CJ, Lopez AD. Evidence-based health policy--lessons from the Global Burden of Disease Study. Science 1996 Nov 1;274(5288):740-743. [Medline: 8966556] [doi: 10.1126/science.274.5288.740]
2. Kristensen T, Kornitzer M, Alfredsson L. Social factors, work, stress and cardiovascular disease prevention in the European Union. Brussels, Belgium: European Heart Network; 1998.
3. Dytell RS, Trent LK, Conway TL. A causal analysis of interrelationships among exercise, physical fitness, and well-being in the U.S. Navy personnel. Report No. 91-27. San Diego, CA: US Naval Health Research Center; 1991.
4. Wielenga RP, Erdman RA, Huisveld IA, Bol E, Dunselman PH, Baselier MR, et al. Effect of exercise training on quality of life in patients with chronic heart failure. J Psychosom Res 1998 Nov;45(5):459-464. [Medline: 9835240] [doi: 10.1016/S0022-3999(97)00309-7]
5. ; National Center for Health Statistics. Health, 2005, United States, with chart book on trends in the health of Americans. Hyattsville, MD: National Center for Health Statistics; 2006.
6. ; US Department of Health and Human Services, Centers for Disease Control and Prevention. Physical activity and health: a report of the Surgeon General . CDC. Atlanta, GA; 1996. URL: http://www.cdc.gov/nccdphp/sgr/pdf/sgrfull.pdf [accessed 2007 Apr 27] [WebCite Cache ID 5OQYBuKvC]
7. ; US Department of Health and Human Services and US Department of Agriculture. Dietary Guidelines for Americans 2005. HealthierUS.gov. URL: http://www.healthierus.gov/dietaryguidelines/dga2005/document/pdf/DGA2005.pdf [accessed 2005 Mar 3] [WebCite Cache ID 5OAdK29Ez]
8. Sallis JF, Bauman A, Pratt M. Environmental and policy interventions to promote physical activity. Am J Prev Med 1998 Nov;15(4):379-397. [Medline: 9838979] [doi: 10.1016/S0749-3797(98)00076-2]
9. Crespo CJ, Keteyian SJ, Heath GW, Sempos CT. Leisure-time physical activity among US adults. Results from the Third National Health and Nutrition Examination Survey. Arch Intern Med 1996 Jan 8;156(1):93-98. [Medline: 8526703] [doi: 10.1001/archinte.156.1.93]
10. Pate RR, Baranowski T, Dowda M, Trost SG. Tracking of physical activity in children. Med Sci Sport Exerc 1996;28(1):92-96. [doi: 10.1097/00005768-199601000-00019]
11. Atkinson NL, Gold RS. The promise and challenge of eHealth interventions. Am J Health Behav 2002;26(6):494-503. [Medline: 12437024]
12. 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]
13. 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]
14. Pinto BM, Friedman R, Marcus BH, Kelley H, Tennstedt S, Gillman MW. Effects of a computer-based, telephone-counseling system on physical activity. Am J Prev Med 2002 Aug;23(2):113-120. [Medline: 12121799] [doi: 10.1016/S0749-3797(02)00441-5]
15. Yoshioka M, Ayabe M, Yahiro T, Higuchi H, Higaki Y, St-Amand J, et al. Long-period accelerometer monitoring shows the role of physical activity in overweight and obesity. Int J Obes (Lond) 2005 May;29(5):502-508. [Medline: 15672105] [doi: 10.1038/sj.ijo.0802891]
16. Tudor-Locke C, Bassett DR, Swartz AM, Strath SJ, Parr BB, Reis JP, et al. A preliminary study of one year of pedometer self-monitoring. Ann Behav Med 2004 Dec;28(3):158-162. [Medline: 15576253] [doi: 10.1207/s15324796abm2803_3]
17. Doshi A, Patrick K, Sallis JF, Calfas K. Evaluation of physical activity web sites for use of behavior change theories. Ann Behav Med 2003;25(2):105-111. [Medline: 12704012] [doi: 10.1207/S15324796ABM2502_06]
18. 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]
19. Hurling R, Fairley BW, Dias MB. Internet-based exercise interventions: are more interactive designs better? Psychol Health 2006;21(6):757-772. [doi: 10.1080/14768320600603257]
20. Festinger L. A theory of social comparison processes. Human Relat 1954;7(2):117-140. [doi: 10.1177/001872675400700202]
21. Janis I, Mann L. Decision making: a psychological analysis of conflict, choice and commitment. New York: Free Press; 1977.
22. Petty RE, Cacioppo JT. The Elaboration Likelihood Model of Persuasion. New York: Academic Press; 1986.
23. Gollwitzer PM. Goal achievement: the role of intentions. Eur Rev of Soc Psychol 1993;4:141-185.
24. Bluetooth.com. URL: http://www.bluetooth.com/bluetooth/ [accessed 2006 Jun 20] [WebCite Cache ID 5OAdOQQC7]
25. ; World Medical Association. Declaration of Helsinki: Ethical principles for medical research involving human subjects. World Medical Association. 1996. URL: http://www.wma.net/e/policy/pdf/17c.pdf

26. ; Canadian Society for Exercise Physiology. PAR-Q & you. CSEP. Ottawa, Ontario: Minister of Public Works and Government Services; 2002. URL: http://www.csep.ca/communities/c574/files/hidden/pdfs/par-q.pdf
27. Cook DG, Shaper AG, MacFarlane PW. Using the WHO (Rose) angina questionnaire in cardiovascular epidemiology. Int J Epidemiol 1989 Sep;18(3):607-613. [Medline: 2807664] [doi: 10.1093/ije/18.3.607]
28. Mathie MJ, Celler BG, Lovell NH, Coster ACF. Classification of basic daily movements using a triaxial accelerometer. Med Biol Eng Comput 2004 Sep;42(5):679-687. [Medline: 15503970] [doi: 10.1007/BF02347551]
29. Bouten CV, Koekkoek KT, Verduin M, Kodde R, Janssen JD. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. IEEE Trans Biomed Eng 1997 Mar;44(3):136-147. [Medline: 9216127] [doi: 10.1109/10.554760]
30. Chen KY, Bassett DR. The technology of accelerometry-based activity monitors: current and future. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S490-S500. [Medline: 16294112]
31. Puyau MR, Adolph AL, Vohra FA, Zakeri I, Butte NF. Prediction of activity energy expenditure using accelerometers in children. Med Sci Sports Exerc 2004 Sep;36(9):1625-1631. [Medline: 15354047]
32. Kumahara H, Schutz Y, Ayabe M, Yoshioka M, Yoshitake Y, Shindo M, et al. The use of uniaxial accelerometry for the assessment of physical-activity-related energy expenditure: a validation study against whole-body indirect calorimetry. Br J Nutr 2004 Feb;91(2):235-243. [Medline: 14756909] [doi: 10.1079/BJN20031033]
33. Cambridge Neurotechnology Ltd home page. URL: http://www.camntech.com/bibl.html [accessed 2007 Apr 3] [WebCite Cache ID 5OAdTTSHs]
34. De Boni M, Hurling R, Hunt DL. Recognizing well-formed implementation intentions in a computer-based health improvement system. In: Proceedings of the IADIS International Conference on Applied Computing. San Sebastian, Spain: IADIS; 2006.
35. Prestwich A, Perugini M, Hurling R. Can text messages strengthen the effects of implementation intentions on exercise? In: BPS Division of Health Psychology Annual Conference; September 13-15. University of Essex, UK: BPS; 2006.
36. De Boni M, Hurling R, Dryden W. A rational-emotive behavior therapy-based automated dialogue system for exercise behavior change. In: Proceedings of the IADIS International Conference WWW/Internet; October 19-22. Lisbon, Portugal: IADIS; 2005.
37. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [Medline: 12900694] [doi: 10.1249/01.MSS.0000078924.61453.FB]
38. Body Composition Analyzer Model BF-350. Tanita Ltd, UK. Tanita Ltd, UK. URL: http://www.tanita.com/ProfessionalManuals/BF-350%20manual.pdf [WebCite Cache ID 5OWGizYSf]
39. Leicester Height Measure. Seca Ltd, UK. URL: http://www.medisave.co.uk/seca-leicester-portable-height-measure-p-1368.html [WebCite Cache ID 5OWQQrrls]
40. Boso Medicus Prestige. Bosch & Sohn, Germany. URL: http://www.boso.de/uploads/media/medic_prest.pdf [WebCite Cache ID 5OWHSP9g3]
41. Whitworth JA, Chalmers J. World health organisation-international society of hypertension (WHO/ISH) hypertension guidelines. Clin Exp Hypertens 2004;26(7-8):747-752. [Medline: 15702630] [doi: 10.1081/CEH-200032152]
42. Williams K, Prevost AT, Griffin S, Hardeman W, Hollingworth W, Spiegelhalter D, et al. The ProActive trial protocol - a randomised controlled trial of the efficacy of a family-based, domiciliary intervention programme to increase physical activity among individuals at high risk of diabetes [ISRCTN61323766]. BMC Public Health 2004 Oct 18;4(1):48 [FREE Full text] [Medline: 15491494] [doi: 10.1186/1471-2458-4-48]
43. Markland D, Tobin V. A Modification to the Behavioural Regulation in Exercise Questionnaire to Include an Assessment of Amotivation. J Sport Exer Psychol 2004;26:191-196.
44. International Physical Activity Questionnaire. Karolinska Institute. URL: http://www.ipaq.ki.se [accessed 2007 Apr 3]
45. Ekelund U, Sardinha LB, Anderssen SA, Harro M, Franks PW, Brage S, et al. Associations between objectively assessed physical activity and indicators of body fatness in 9- to 10-y-old European children: a population-based study from 4 distinct regions in Europe (the European Youth Heart Study). Am J Clin Nutr 2004 Sep;80(3):584-590 [FREE Full text] [Medline: 15321796]
46. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Montoye HJ, Sallis JF, et al. Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc 1993 Jan;25(1):71-80. [Medline: 8292105] [doi: 10.1249/00005768-199301000-00011]
47. Ainsworth BE, Strath SJ, Swartz AM, O'Brien WL, Thompson RW, Jones DA, et al; Bassett DRJr. Comparison of three methods for measuring the time spent in physical activity. Med Sci Sports Exerc 2000;32(9).
48. SAS version 9.1.3. SAS Institute. Cary, NC, USA. .
49. Miller WC, Eggert KE, Wallace JP, Lindeman AK, Jastremski C. Successful weight loss in a self-taught, self-administered program. Int J Sports Med 1993 Oct;14(7):401-405. [Medline: 8244607]
50. Klem ML, Klesges RC. Competition in a minimum-contact weight loss program. J Consult Clin Psychol 1998;56:142-144.
51. Franks P, Gold MR, Fiscella K. Sociodemographics, self-rated health, and mortality in the US. Soc Sci Med 2003 Jun;56(12):2505-2514. [Medline: 12742613] [doi: 10.1016/S0277-9536(02)00281-2]

52. Brown DW, Balluz LS, Heath GW, Moriarty DG, Ford ES, Giles WH, et al. Associations between recommended levels of physical activity and health-related quality of life. Findings from the 2001 Behavioral Risk Factor Surveillance System (BRFSS) survey. Prev Med 2003 Nov;37(5):520-528. [Medline: 14572437] [doi: 10.1016/S0091-7435(03)00179-8]
53. Fogg BJ. Persuasive Technology. San Francisco: Morgan Kaufmann; 2002.
54. Womble LG, Wadden TA, Mcguckin BG, Sargent SL, Rothman RA, Krauthamer-ewing ES. A randomized controlled trial of a commercial internet weight loss program. Obes Res 2004 Jun;12(6):1011-1018 [FREE Full text] [Medline: 15229342]
55. Plasqui G, Joosen AMCP, Kester AD, Goris AHC, Westerterp KR. Measuring free-living energy expenditure and physical activity with triaxial accelerometry. Obes Res 2005 Aug;13(8):1363-1369 [FREE Full text] [Medline: 16129718]
56. Williams P, Fitzsimons GJ, Block G. When consumers do not recognize benign intention questions as persuasion attempts. J Consum Res 2004;31(3):540-550. [doi: 10.1086/425088]
57. Ajzen I. The theory of planned behavior. Organ Behav Human Decis Process 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
58. Conner M, Sparks P. Theory of planned behavior and health behavior. In: Conner M, Norman P, editors. Predicting Health Behavior, p. 121-162. Buckingham, UK: Open University Press; 1996.
59. Prestwich AJ, Lawton RJ, Conner MT. The use of implementation intentions and the decision balance sheet in promoting exercise behaviour. Psychol Health 2003;18(6):707-721. [doi: 10.1080/08870440310001594493]
60. Prestwich AJ. Motivational and volitional interventions in exercise promotion: is two better than one? Health Psychol Update 2003;12(2):13-15.
61. Hillsdon M, Thorogood M. A systematic review of physical activity promotion strategies. Br J Sports Med 1996 Jun;30(2):84-89. [Medline: 8799589]
62. Ronda G, Van Assema P, Brug J. Stages of change, psychological factors and awareness of physical activity levels in The Netherlands. Health Promot Int 2001 Dec;16(4):305-314 [FREE Full text] [Medline: 11733449] [doi: 10.1093/heapro/16.4.305]
63. Conn VS, Valentine JC, Cooper HM. Interventions to increase physical activity among aging adults: a meta-analysis. Ann Behav Med 2002;24(3):190-200. [Medline: 12173676] [doi: 10.1207/S15324796ABM2403_04]
64. Blair SN. Methods for voluntary weight loss and control: national institutes of health technology assessment conference: evidence for success of exercise in weight loss and control. Ann Intern Med 1993;119:702-706.
65. Ross R, Janssen I, Dawson J, Kungl AM, Kuk JL, Wong SL, et al. Exercise-induced reduction in obesity and insulin resistance in women: a randomized controlled trial. Obes Res 2004 May;12(5):789-798 [FREE Full text] [Medline: 15166299]
66. Garrow JS, Summerbell CD. Meta-analysis: effect of exercise, with or without dieting, on the body composition of overweight subjects. Eur J Clin Nutr 1995 Jan;49(1):1-10. [Medline: 7713045]
67. Sechrist KR, Walker SN, Pender NJ. Development and psychometric evaluation of the exercise benefits/barriers scale. Res Nurs Health 1987 Dec;10(6):357-365. [Medline: 3423307] [doi: 10.1002/nur.4770100603]
68. Cassell MM, Jackson C, Cheuvront B. Health communication on the Internet: an effective channel for health behavior change? J Health Commun 1998;3(1):71-79. [Medline: 10947375] [doi: 10.1080/108107398127517]
69. Skinner BF. Contingencies of reinforcement: a theoretical analysis. New York: Appleton-Century-Crofts; 1969.
70. Dishman RK, Buckworth J. Increasing physical activity: a quantitative synthesis. Med Sci Sports Exerc 1996 Jun;28(6):706-719. [Medline: 8784759]

Abbreviations

BMI: body mass index
IPAQ: International Physical Activity Questionnaire
MET: metabolic equivalent

submitted 18.02.07; peer-reviewed by A Dey, S Johnston, L Barkhuus; comments to author 25.02.07; revised version received 16.03.07; accepted 30.03.07; published 27.04.07.

Please cite as:
Hurling R, Catt M, De Boni M, Fairley BW, Hurst T, Murray P, Richardson A, Sodhi JS
Using Internet and Mobile Phone Technology to Deliver an Automated Physical Activity Program: Randomized Controlled Trial
J Med Internet Res 2007;9(2):e7
URL: http://www.jmir.org/2007/2/e7/
doi:10.2196/jmir.9.2.e7
PMID:17478409

Original Paper

A Survey of Quality Assurance Practices in Biomedical Open Source Software Projects

Günes Koru[1], PhD; Khaled El Emam[2], PhD; Angelica Neisa[3]; Medha Umarji[1]

[1]Department of Information Systems, University of Maryland, Baltimore, MD, USA

[2]University of Ottawa and Children's Hospital of Eastern Ontario (CHEO) Research Institute, Ottawa, ON, Canada

[3]Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada

Corresponding Author:
Khaled El Emam, PhD
CHEO Research Institute
401 Smyth Rd
Ottawa, ON K1H 8L1
Canada
Phone: +1 613 738 4181
Fax: +1 613 741 0397
Email: kelemam@uottawa.ca

Abstract

Background: Open source (OS) software is continuously gaining recognition and use in the biomedical domain, for example, in health informatics and bioinformatics.

Objectives: Given the mission critical nature of applications in this domain and their potential impact on patient safety, it is important to understand to what degree and how effectively biomedical OS developers perform standard quality assurance (QA) activities such as peer reviews and testing. This would allow the users of biomedical OS software to better understand the quality risks, if any, and the developers to identify process improvement opportunities to produce higher quality software.

Methods: A survey of developers working on biomedical OS projects was conducted to examine the QA activities that are performed. We took a descriptive approach to summarize the implementation of QA activities and then examined some of the factors that may be related to the implementation of such practices.

Results: Our descriptive results show that 63% (95% CI, 54-72) of projects did not include peer reviews in their development process, while 82% (95% CI, 75-89) did include testing. Approximately 74% (95% CI, 67-81) of developers did not have a background in computing, 80% (95% CI, 74-87) were paid for their contributions to the project, and 52% (95% CI, 43-60) had PhDs. A multivariate logistic regression model to predict the implementation of peer reviews was not significant (likelihood ratio test = 16.86, 9 df, P = .051) and neither was a model to predict the implementation of testing (likelihood ratio test = 3.34, 9 df, P = .95).

Conclusions: Less attention is paid to peer review than testing. However, the former is a complementary, and necessary, QA practice rather than an alternative. Therefore, one can argue that there are quality risks, at least at this point in time, in transitioning biomedical OS software into any critical settings that may have operational, financial, or safety implications. Developers of biomedical OS applications should invest more effort in implementing systemic peer review practices throughout the development and maintenance processes.

(J Med Internet Res 2007;9(2):e8) doi:10.2196/jmir.9.2.e8

KEYWORDS

Open source software; medical informatics; computational biology; information systems; software quality assurance; software/program verification; code inspections and walkthroughs; software reliability

Introduction

The importance of Information and Communications Technology (ICT) to the health care industry is rising as organizations attempt to find ways of reducing the costs of care and improving patient safety. However, in general, ICT adoption in health care is underfunded. For example, in Canada, the proportion of the 2003 national health care budget devoted to ICT was approximately 1.8% [1]. Canadian hospitals spend 1.8% to 2.5% of their budgets on ICT, which is low compared

with other nations [2]. Recently ICT expenditures in Canadian hospitals have even been decreasing [3]. Other information-intensive sectors, such as banking and government, have ICT expenditures ranging from 9% to 13% of their operating budgets [1]. In the United States, the health care industry invests only 2% of gross revenue in ICT [4].

Therefore, cost is an important barrier to the adoption of information systems in health care [4,5]. This is motivating increased interest in open source (OS) software [6,7]. Surveys show that enterprises adopt OS primarily because they believe it is cheaper to acquire than the alternative solutions and believe it to have a lower total cost of ownership [21-24]. Many information technology (IT) managers also believe that OS is free [12]. It has been suggested that OS can address some of the biggest barriers to the adoption of electronic medical records (EMRs) by physicians [7,13]. Medicare in the United States has made a modified version of the OS Vista system freely available to all US medical practices [14]. In general, there is an increasing uptake of OS EMRs [15]. There also exist OS tools to support clinical trials [16] and imaging tools for radiologists [17]. In the bioinformatics domain, OS software has already gained popularity [21-24].

Software used in providing care is often safety and mission critical. There is evidence that software failures in clinical information systems and in medical devices have caused serious injuries and fatalities in patient populations [21-24]. Research and clinical applications are merging [25], highlighting the quality risks for research applications as well.

Ensuring that software, including OS software, is of high quality and that the processes used to develop that software follow best software engineering practice is becoming a necessity. However, it has been suggested that the OS development model used for many bioinformatics products can have weaknesses in terms of quality assurance (QA) compared to the development of non-OS products that typically involve formal QA procedures [25].

In the commercial world, third-party process assessments (such as ISO 9000:2000 or Capability Maturity Model Integration [CMMI] [26]) are used to provide some assurance to end users that effective software engineering practices are in place during software development, in particular, QA practices. In an OS project, it is not possible to perform such assessments.

To facilitate the adoption of OS in the biomedical domain, and to understand the risks, if any, this paper reports on a Web survey of the QA practices used by the global biomedical OS developer community. The objectives of the survey were twofold:

1. To determine the rate at which two key QA practices, peer reviews and testing, are actually being used in biomedical OS projects
2. To determine whether the following factors have an impact on the implementation of these practices: experience of the developers, educational background of developers, size of the software product, and number of users

To our knowledge, this is the first study to focus on the QA practices of OS developers in the biomedical domain.

Literature on QA Practices in OS Development

The current evidence on the QA practices implemented in general (nonbiomedical) OS projects was summarized. The two main QA activities we focused on were peer reviews and testing because they are the most commonly practiced in software development projects.

A literature search was conducted in the fall of 2005, primarily on EI Compendex and Inspec databases. The search was restricted to English articles published since 1995. To ensure that we cast a wide net, we used the term “open source software” for the search. The ACM digital library, IEEE Xplore, and SpringerLink were also searched. The reference lists from included studies were examined in an effort to identify additional articles and relevant dissertations.

A total of 2731 articles were initially identified. A level one screening based on the titles and abstracts removed articles that did not discuss software quality practices directly or indirectly. The second level of screening based on the full text included only articles that had original data on QA practices.

The quality of the studies varied greatly. There were no controlled experiments, and many of the original data came from surveys or qualitative studies. Below we summarize the evidence descriptively and then draw conclusions based on the consistency of evidence from multiple sources and using multiple methodologies.

Peer-Review Practices

One of the fundamental assumptions for the success of OS is that there will be many developers looking over the source code. The popularized statement “given enough eyeballs, all bugs are shallow” captures that assumption [27]. The argument goes that because so many developers and users look at the source code of an OS application, this amounts to a large-scale peer-review process. Some authors are talking about millions of programmers performing peer review on OS code [28].

It is well established in software engineering that peer reviews are one of the best methods for defect detection [29]. In addition, in non-OS settings, the prospect of their code being scrutinized by their peers motivates programmers to be more careful [30,31].

Peer reviews are a relatively common practice in OS projects. For example, in the FreeBSD project, 57% of the respondents working on the project had distributed code for review in the month prior to being surveyed, and 86% of those who distribute code for review do get feedback from their peers [31]. In another survey, only 9% of OS developers indicated that all of their code was peer reviewed, but almost 50% indicated that most of their code was reviewed [32].

How does this compare to non-OS projects? One survey found that 52% of companies perform code reviews [33]. Another survey found that 71% of respondents in the United States use code reviews, around 82% in Europe, and 74% in Japan [34]. Therefore, there is evidence that a majority of non-OS software project developers also use some form of peer review.

In one study, the average number of reviewers per OS application was 1.3, including the author, and for larger projects,

this number climbs to 7.5 reviewers [35]. The number of reviewers on larger OS projects is consistent with the numbers typically recommended in the software engineering literature for commercial (non-OS) projects. Porter et al [36] reported that peer reviews are usually carried out by a team of 4 to 6 inspectors. The recommended size for reviews defined by the Institute of Electrical and Electronics Engineers (IEEE) standard for software reviews and audits is 3 to 6 persons [37]. Ackerman et al [38] reported that reviews are conducted by at least 3 people, one of whom is the moderator who is responsible for the effectiveness of the examination. Lau et al [39] suggested 3-person reviews. Grady [40] stated an optimum size of 4 to 5 reviewers. Laitenberger and DeBaud [41] recommended 3 to 4 reviewers. They stated that there would be a ceiling effect after which an additional reviewer would not necessarily pay off in more defects being discovered. Fagan [42] recommended keeping review teams small, that is, 4 people. Weller [43] reported that 4-person teams were better at finding bugs than 3-person teams. Beyond 4 members, there is no performance improvement. Madachy et al [44] suggested an optimal size of 4 to 5 people for reviews. Gilb and Graham [45] mentioned a team size of 4 to 5 people to maximize the ability to find bugs. Strauss and Ebeneau [46] suggested a minimum review team size of 3 and a maximum of 7. In an experiment at the Jet Propulsion Laboratory, Kelly et al [47] stated that reviews are usually carried out by 6 people. Johnson [48] notes that there is widespread consensus that the review team size should never exceed 6 to 9 members. Industrial practice varies even within a single enterprise, for example, review teams between 4 and 12 are used at AT&T [49].

Two relevant characterizations of peer reviews are “ad-hoc” and “checklist” peer reviews. The former means that the reviewers do their best to find defects without any guidance. The latter means that a standardized checklist of common defects is used to guide the reviewers and to ensure that they look for all defect types. There is evidence that checklist-based techniques tend to find more defects than ad-hoc techniques [50]. There have been no studies on whether OS peer reviews utilize ad-hoc or checklist approaches.

Testing Practices

Most OS projects do not perform extensive pre-release testing internally. The most common testing performed is unit testing. After that, a release candidate is made available, and the users try the release candidate and report its defects.

Below is a summary of the results of studies on pre-release testing activities and how well they are at finding bugs.

In general, pre-release testing is not a prevalent activity [51] and formal testing is not common [52]. The prospect of having their code peer reviewed makes OS developers more comfortable releasing their software with little testing [32], and more than 80% of OS developers in another survey responded that their products do not have testing plans [35]. This is consistent with other evidence showing that only 32% of projects had design documents [53] and that OS projects typically do not produce explicit requirements [54,55], making testing against requirements and design more difficult.

A majority of OS developers believe that anyone who downloads their code will check for and report bugs (implying that if there are no bug reports then no one has found anything) [35]. This indicates that testing effort by developers is minimal as they tend to rely on other people to look for defects.

The evidence on automated tool usage is mixed. A common testing tool is a debugger [35]. Only 48% of OS projects used a baseline test suite to support regression testing, and the percentage is only slightly higher as projects become large [53]. One early study reported that the Apache project had no regression or system test performed [54]. A more recent analysis did indicate that there was a regression test suite for Apache, but its use was not mandatory [56]. The Linux kernel did not go through pre-release testing by the developers, but rather users report defects in release candidates [56]. With the recent commercialization of Linux, considerable effort has been put into producing regression test suites [57], although an analysis of the test coverage of some of these test suites found that it was rather poor, with many critical subsystems having low or no coverage [58]. Subversion, another OS project, had an automated regression test suite [56]. Mozilla had dedicated test teams and test plans [54].

If we compare the above numbers with recent data on the implementation of regression testing in commercial software development, one can see that OS projects are somewhat lagging [34].

More than 50% of surveyed OS projects did not take advantage of code coverage metrics [35]. Only 5% of projects employed tools to measure test coverage [53], and almost 30% of projects had an (subjectively) estimated test coverage of less than 30% [53].

The above evidence indicates that developer pre-release testing in many OS projects is not a priority.

Peer Reviews vs Testing

The argument has been made that the lack of developer pre-release testing is compensated for by the peer reviews that are conducted [52]. The problem in this argument is that there is substantial evidence that peer review will not find the same types of bugs (probabilistically) that testing will find [59-63]. Therefore, peer reviews and testing are not alternatives but rather are complementary. Consequently, doing only peer reviews will likely result in a smaller proportion of bugs being discovered than employing both peer reviews and testing.

One of the arguments made in support of OS quality is that OS programs tend to have a large number of end users who effectively beta test the software every time it is released by going into the code to identify where the bugs are and contributing the patches to correct these bugs. Such large-scale and free beta testing and debugging ensures that any defects that escape development will be discovered relatively quickly. The argument has some merit: it is known that defect discovery is affected by usage. Therefore, the more users a product has, the more defects will be discovered. However, as indicated below, few of the users actually contribute bug reports, and fewer still contribute bug fixes. Therefore, the number of users

is not an appropriate reflection of the amount of debugging activity that is going on.

A study of the Apache project by the Software Engineering Institute [64] found that the majority (93%) of the changes (implementations, patches, and enhancements) were made by the core group of developers. The total number of people reporting bugs was 5116, but only 200 individuals actually contributed patches. The difficult and critical architectural changes were made by an even smaller subset of the core developer group [64].

Another investigation of Apache [65] found that more than 83% of the modification requests came from the top 15 developers, as did 88% of added lines of code and 91% of deleted lines of code. About 66% of bug fixes were produced by the top 15 developers, but 182 individuals submitted bug fixes out of 3060 who submitted bug reports [54]. The top 15 problem reporters submitted only 5% of the problem reports. These numbers indicate that new functionality is developed by a small core team, but there is wider participation in bug fixes. However, most of the people who report bugs do not actually submit any fixes. A small proportion of those who report problems are truly debuggers.

An interesting informal survey that was performed among Unix users (comprising researchers, staff, and students) at a computer science department asked if users had encountered bugs and if they reported them [66]. All of the respondents who had encountered bugs, some serious, did not bother to report them. To the extent that this behavior is common among technical users (Unix users tend to be rather technically savvy), many users will not bother reporting bugs even if they do discover them. In the case of Apache, it was estimated that less than 1% of all Apache users report problems [54].

In addition, most (80%) of the OS projects have less than 11 end users (where subscribers is used as a surrogate for users) [67]. Only 1% have more than 100 users [67]. A Pareto analysis of active projects on SourceForge found that half of the active projects had between 0 and 70 downloads per month [68] and that a very small number of projects are popular, with the vast majority not experiencing many downloads. Following a Pareto distribution means that the number of projects with more than a given number of downloads tails off exponentially.

Therefore, we can conclude that for most OS projects the number of users tends to be relatively small, the proportion of these users who report bugs is smaller, and the proportion of those who contribute patches are even smaller. It can be argued, then, that the extent of community debugging and beta testing is not that extensive in practice.

Literature Review Summary

The existing evidence paints a decidedly mixed picture of the QA practices of OS projects. One can conclude that the better projects have practices that are, at best, comparable with non-OS projects. In general, peer reviews and testing, when practiced, tend to be minimal. Such practices are consistent with the results from studies directly measuring the postrelease quality of OS applications. An evaluation of the postrelease defect levels in Apache found that defect density was higher than a number of commercial products in the telecommunications field [54,65]. A recent study of FreeBSD also collected postrelease defect data [69], with mixed results when compared to non-OS products.

None of the studies that we found focused on biomedical applications—we do not know what the QA practices are for biomedical OS projects or what the resulting quality is. Therefore, it is unclear whether the conclusions from general OS studies can be extended to the biomedical domain.

Methods

Questionnaire Development

Our Web questionnaire was based on two previous general (nonbiomedical) OS developer surveys. The first survey was conducted by Stark [32,70] to understand the peer-review practices in OS projects. The second was performed by Zhao and Elbaum [35,53] and was geared toward understanding the QA activities in OS projects.

There were five sections in the questionnaire relevant to this paper: (1) respondent demographics, (2) project characteristics, (3) extent and nature of the implementation of peer reviews, (4) extent and nature of the implementation of testing, and (5) optional contact information (Multimedia Appendix 1).

A pilot study was conducted with software development staff at Georgetown Medical Center and the National Cancer Institute. A draft of the online questionnaire was sent to the two pilot sites, and comments were solicited on ease of understanding the questionnaire, the usability of the Web form, and the time it takes to complete. The questionnaire was revised based on this feedback.

Survey Setup

The target population for the survey consisted of developers of biomedical OS projects. A project was considered an OS project only if its source code was publicly available. We made a list of biomedical OS projects by searching for OS projects in bioinformatics, medical informatics, and health care informatics domains. Biomedical OS projects were selected based on a Web search and expert inputs. The initial list was constructed based on the authors' knowledge of OS applications. We then went to the OS project-hosting websites SourceForge [71] and FreshMeat [72] and identified additional projects. On SourceForge, we identified the projects listed under bioinformatics and medical science applications subcategories, which are under the Scientific/Engineering main project category. On FreshMeat, there was the exact same categorization, so we identified projects in the same way. We also used the BioMed Central website [73], where software developed for various biomedical research projects was available. Finally, we sent the project list to our colleagues and asked whether they would add any OS projects to our list. While creating the list of projects, we encountered some projects hosted on multiple sites with the same or similar names. Such duplicate project entries were eliminated. As a result of this in-depth search, we identified 229 projects (Multimedia Appendix 2).

After obtaining the list of projects, we started to identify the names and email addresses of the developers working on those projects using the following information sources:

1. Project websites: The names and email addresses of some developers were listed on project websites.
2. Source code repositories: OS projects often adopt a configuration management system, for example, Concurrent Versions Systems (CVS) or Subversion to allow developers to manage different versions of their source code files. Developers check in and check out source files to and from this repository. Usually, a log is kept in the source code repository for each checkin. The log includes the developer's log-in name and email address. From these logs, we were able to identify some developers.
3. Defect databases: OS projects usually employ a defect-handling tool, Bugzilla, or a variant of it [74]. The defect records in these databases include rich information, such as the log-in names and email addresses of the developers assigned to solve the defects, which can be extracted.
4. Mailing lists: Developers in OS projects usually communicate using a designated mailing list. The emails sent to these lists are archived on the project websites. The header portions of the emails include the names and email addresses of developers.

Using multiple information sources allowed us to cross-validate the lists. Duplicate developer entries were eliminated. In our developer list, we did not include those who made minor contributions to the projects by occasionally fixing bugs, sending emails, or committing source code. As a result, our sample consisted of 750 developers heavily involved in the targeted projects. A small number of developers were involved in more than one project. We made it clear to those developers that they should base their questionnaire answers on a single project that we selected. Therefore, many developers from a project were allowed, but one developer could only answer for one project.

During and after the survey period, we took extensive precautions to maintain the confidentiality of the respondents' records. Other than the one described above, no other prescreening or identification procedure was used.

The identified OS developers were sent an invitation email and a link to the final Web survey. After the first week, multiple reminders were sent to nonrespondents over a period of 6 weeks.

Analysis Methods

The analysis was performed at two levels, the individual and project level. This suggests a multi-level approach to analysis. However, there were structural reasons why such a hierarchical modeling approach was not deemed appropriate in this case: most OS projects are small. In general, at least 5 respondents per project are recommended in order to model multiple levels and their interactions [75,76]. Of the 229 projects that were surveyed, only 23 (10%) had more than 5 developers. Of the 106 projects that we received responses for, 20 (19%) had more than one developer and only 2 (~2%) had more than 5 developers. Therefore, an alternative approach was necessary.

When reporting individual-level results, we will use all of the respondents' records. For project-level analysis, individual responses in each project were aggregated. In the respondent database there were 138 observations and 106 projects. For aggregation, the most experienced developer's (as determined by responses to the demographic questions) response was selected to represent the values for the project.

To address the first objective of the study (extent of use), descriptive statistics on the extent of use of the two QA practices were reported as proportions (percentages) with 95% confidence intervals.

To address the second objective (factors affecting the extent of use), multivariate logistic regression models [77] were developed for each of the main outcomes being investigated: implementation of peer reviews, measured by Q11 of the survey, and implementation of testing, measured by Q21 of the survey (see Multimedia Appendix 1). The unit of analysis was the project. The predictors consisted of the developer demographics and the project characteristics: years of programming experience (Q1 and Q2), whether the developer had a computing background (Q3), the number of users of the product (Q6), and the size of the product (Q9). It is reasonable to expect that the more experienced the developers, the more likely they will implement better software engineering practices. Also, we assumed that developers with a stronger computing background would be more likely to be associated with the implementation of key QA practices. The more users of the product, then the more individuals who are available to peer review, and this has been one of the core arguments made in support of OS software [27]. Finally, larger projects require the development team to impose more discipline and better practices to ensure the sustainability of the development effort (so that the project does not descend into a continuous cycle of bug fixes, each introducing even more bugs).

Response Rate and Nonresponse Bias

There were no missing data since the Web survey tool made all questions mandatory. Therefore, all submitted forms were complete.

Since we performed our analysis separately at the individual and project level, we report the response rates for both. We received responses from 106 of the 229 projects contacted, which gave us a project-level response rate of 46.3%. Out of 750 developers contacted, 138 of them replied, which corresponded to a response rate of 18.4% at the individual level. Other Web surveys have shown a comparable response rate at the individual level [78]. The response rates in our survey were also comparable to previous surveys of OS developers, which ranged from 21% [79,80] to 34% [81].

Nonresponse bias was evaluated by comparing early respondents with late respondents [82]. We took the respondents who replied before the first reminder as early respondents. We compared the demographics of the early/late respondents and the characteristics of their projects using the Wilcoxon rank sum test [83]. None of the differences were significant at the .05 alpha level, indicating that the early and late respondents were identical in background characteristics.

Summary

A summary of the survey setup and administration according to the CHERRIES guidelines [84] is provided in Multimedia Appendix 3.

Results

Background of Biomedical OS Developers

As can be seen from Table 1, 73% of the respondents (95% CI, 66,-81) had at least 5 years' experience writing software. A large percentage of that experience was in the biomedical area. Therefore, one would expect these developers, in general, to have a good understanding of the computational needs in that domain.

In terms of project participation, 71% (95% CI, 63-79) of the respondents were participating in their projects part-time. Contrary to the assumption that OS developers do not receive compensation for their efforts, about half of all respondents were part-time and were paid by their employers for their contributions, and 80% (95% CI, 74-87) of the respondents received either part-time or full-time support for their development effort.

We also looked at the highest attained degree of the respondents, as summarized in Table 1. The biomedical OS developers were qualified professionals in their domain with 52% (95% CI, 43-60) of them having PhDs. A differentiation is made between those who had a computer science (or computer engineering) degree and those who did not (eg, biology, genetics, biochemistry, and physics). This distinction was based on the assumption that the computer science and computer engineering graduates would have a stronger grounding in software engineering practices than graduates of other disciplines. Almost three quarters of respondents (74%; 95% CI, 67-81) did not have a computing background.

Table 1. Developer education and experience (n = 138)

	%	No.
Years of programming experience		
< 1 year	2	3
1-5 years	25	34
> 5 years	73	101
Years of experience in developing biomedical software		
< 1 year	4	5
1-5 years	53	73
> 5 years	43	60
Project participation level		
Part-time, supported by employer	51	71
Part-time, personal time	20	27
Dedicated, full-time	29	40
Highest academic degree and subject area		
Bachelors in CS/CE	7	10
Masters in CS/CE	12	17
PhD in CS/CE	7	9
Bachelors in non-CS/CE	10	14
Masters in non-CS/CE	16	22
PhD in non-CS/CE	45	62
MD	3	4

CS/CE = computer science or computer engineering

Table 2 shows the experience level of the respondents in peer reviewing others' code and in testing. Around 28% (95% CI, 21-36] of the developers had never peer reviewed others' code, and approximately 19% (95% CI, 12-25) of them had received formal education in testing.

Table 2. Quality assurance experience of developers (n = 138)

	%	No.
Years of experience in peer reviewing others' code		
None	28	39
< 1 year	12	17
1-5 years	28	38
> 5 years	32	44
Formal education in testing		
No	81	112
Yes	19	26

Product and Project Characteristics

We found that 50% (95% CI, 40-60) of the products had at least 50 users. Therefore, it can be said that the quality of these biomedical products affects a large group of users. Approximately 63% (95% CI, 54-72) of the products are released at 6-month intervals or less, which is quite a rapid release cycle. Just under one third (30%; 95% CI, 21-39) of the products had been available for more than 3 years. Around a quarter (27%; 95% CI, 19-36) were larger than 50000 lines of code in size.

Peer-Review Practices

Peer review is widely accepted to be one of the important strengths of the OS development model. However, for 63% (95% CI, 54-72) of the projects, peer review was not made an integral part of the development process, and peer review was never performed for 40% (95% CI, 30-49) of the projects.

Table 3 shows results for projects that did perform peer reviews (n = 64). We found that, in 81% (95% CI, 74-89) of those projects, peer reviews were never or only occasionally performed before the code is committed, and in 64% (95% CI, 55-73) of those projects, peer reviews were never or only occasionally performed before the product is released. A majority of projects (84%; 95% CI, 77-91) did not use checklists during their peer-review activities.

Table 3. Peer review practices at the project level, for projects that did perform some peer review (n = 64)

	%	No.
Source code is _____ peer reviewed before commit.		
Never	12	8
Occasionally	69	44
Half the time	3	2
Frequently	8	5
Almost always	8	5
Source code is _____ peer reviewed before product release.		
Never	6	4
Occasionally	58	37
Half the time	6	4
Frequently	10	6
Almost always	20	13
Do you use a checklist for peer review?		
No	84	54
Yes	16	10

Almost two fifths of all respondents (40%; 95% CI, 32-48) said that they never reviewed someone else's code, 36% (95% CI, 28-44) never asked someone else to review their code, and 41% (95% CI, 32-49) said that no one reviewed their code. Of the survey respondents who asked others to review their code, 93% stated that three or fewer individuals do the review.

For those projects that did not perform peer-reviews (n = 42) we also asked for the reasons why. A high proportion (40%; 95% CI, 31-50) did not perform peer-reviews because there

were other things to do ("work is too busy"), and 12% (95% CI, 6-18) were not because the developers believed that the code was already of sufficiently high quality that peer reviews were not needed. 17% were unsure how to review, and 7% said reviewing brings no beneft. For some projects, under the "Other" option, it was stated that the code was so large that it was not possible to apply peer reviews. In some projects, developers stated that they were the only developer, and they did not consider asking others to do peer reviews because they thought no one else would understand their code.

Testing Practices

Testing was an integral part of the development process for 82% (95% CI, 75-89) of projects, and a regression test suite was run before every release for 58% (95% CI, 48-67) of the projects. The percentage of the projects for which a baseline test suite was used was 56% (95% CI, 46-65). Automated tools were used during development for only 25% (95% CI, 16-33] of the projects. One would conclude that automated testing is done after the development work is complete. Only 4% (95% CI, 0-7) of projects had automated test coverage tools. For the 21% (95% CI, 13-28) of the projects, the developer selected the "don't know" option when asked about estimated code coverage. Only around 25% of the projects exceeded 80% code coverage.

As shown in Table 4, different types of testing are used in the projects. Unit testing is performed in 78% (95% CI, 70-86) of the projects. Approximately 70% of the projects (95% CI, 61-79) had unit testing performed half the time or more. Integration and system testing are also quite common. Testing is conducted continuously in 60% (95% CI, 51-70) of projects. In 32% (95% CI, 23-41) of projects, testing continues after releasing the software to specific users. Most defects that are fixed are found through testing rather than being discovered through usage.

Table 4. Testing practices and test results (n = 106)

	%	No.
Do you perform _____?		
Unit testing	78	83
Integration testing	64	68
System testing	75	80
System load and performance testing	45	48
Other	4	4
How often do you unit test?		
Never	8	9
Occasionally	22	23
Half the time	8	8
Frequently	17	18
Almost always	45	48
What percentage of fixed defects is discovered by testing?		
< 20%	16	17
20-40%	21	22
40-60%	23	24
60-80%	25	27
> 80%	15	16
What percentage of fixed defects is discovered by users?		
< 20%	47	50
20-40%	20	22
40-60%	15	16
60-80%	9	9
> 80%	9	9
Testing is performed _____.		
Continuously	60	64
Before release	57	60
After releasing to specific users	32	34
Randomly	29	31
Other	4	4

In Table 5, it can be seen that 80% (95% CI, 74-87) of the respondents spent less than 40% of their time on testing. The most common practices to generate test cases were the imitation of valid user behavior, generating failure inducing inputs, and using personal experience. Random testing and extreme load testing were not common.

Table 5. Testing practices at the individual level (n = 138)

	%	No.
What percentage of the development time is spent on testing?		
< 20%	44	61
20-40%	36	50
40-60%	14	19
60-80%	4	5
> 80%	2	3
What strategies are adopted in choosing the test cases?		
Provide inputs to imitate valid user behavior	82	113
Choose inputs most likely to cause failures	67	92
Choose inputs according to your experience	72	100
Use scripts to provide random inputs	20	28
Provide extreme values as inputs	46	63
Provide boundary conditions as inputs	43	59
Try extreme loads	28	38

Multivariate Models

The objective of the multivariate models was to understand the factors that have an impact on the implementation of peer reviews and testing in biomedical OS projects. An initial analysis indicated that the two measures of years of programming experience (see Q1 and Q2 in Multimedia Appendix 1) were strongly correlated. We therefore constructed our models using Q1 only. The logistic regression model for the implementation of peer reviews was:

logit(Q10) ~ Q1 + Q3 + Q6 + Q9

And the logistic regression model for the implementation of testing was:

logit(Q21) ~ Q1 + Q3 + Q6 + Q9

The model variables are the answers to the following questions:

- Q1: The number of years of programming experience the respondent has.
- Q3: Whether the highest academic degree obtained by the respondent was in a computer science or a related area.
- Q6: The estimated current number of users of the OS product.
- Q9: The approximate size of the OS product.
- Q10: Whether peer review is an integral part of the project's software development process.
- Q21: Whether testing was an integral part of the project's software development process.

All independent variables except Q9 were ordinal. Therefore, repeated contrasts coding [85,86] was used to capture this ordering. For an independent variable with k ordered categories, $k - 1$ independent coding variables are used. The parameters for these coding variables represent the change in logit when the independent variable changes from one category to the next category.

The overall results for the main effect models are shown in Table 6, and the detailed parameter estimates in Table 7. An alpha level of .05 was used for all tests. These results indicate that none of main effect models are an improvement over the null model (with the intercept only).

An analysis of deviance comparing models with interaction effects versus main effects indicated that there were no interaction effects.

Table 6. The main effect models for the two outcome variables

Model	Likelihood Ratio Test	Nagelkerke R^2
Peer review	16.84 (9 df); P = .051	0.201
Testing	3.34 (9 df); P = .95	0.051

Table 7. Detailed model parameter estimates for the two logistic regression models

Variable	Beta Coefficient*	Odds Ratio*	95% CI*	*P* value†
Peer Review Model				
Intercept	8.76	–	–	.74
Q1: Programming experience (years)				
> 1-5 vs < 1	−0.15	0.86	−0.13 to 1.85	.80
5+ vs 1-5	−7.80	0.00	−0.02 to 0.02	.77
Q6: Number of users				
5-10 vs < 5	1.03	2.81	−0.5 to 6.11	.08
10-50 vs 5-10	0.014	1.01	−0.51 to 2.54	.98
50+ vs 10-50	−1.62	0.20	−0.17 to 0.57	.09
Q9: Size (lines of code)				
5000-20000 vs < 5000	−0.49	0.61	−0.15 to 1.38	.44
20000-50000 vs 5000-20000	1.21	3.37	−1.11 to 7.85	.07
> 50000 vs 20000-50000	−1.41	0.24	−0.07 to 0.56	.03
Q3: CS degree	0.33	1.39	0.05 to 2.73	.50
Testing Model				
Intercept	0.51	–	–	.79
Q1: Programming experience (years)				
1-5 vs < 1	−0.77	0.46	−0.3 to 1.23	.36
5+ vs 1-5	2.19	5.39	−8.35 to 19.13	.19
Q6: Number of users				
5-10 vs < 5	−0.17	0.84	−0.32 to 2.00	.81
10-50 vs 5-1-10	0.011	1.01	−0.67 to 2.70	.99
50+ vs 10-50	−0.33	0.72	−1.06 to 2.51	.79
Q9: Size (lines of code)				
5000-20000 vs < 5000	−0.21	0.81	−0.47 to 2.09	.79
20000-50000 vs 5000-20000	0.20	2.23	1.09 to 3.36	.80
> 50000 vs 20000-50000	−0.03	0.96	−0.44 to 2.37	.96
Q3: CS degree	−0.39	0.68	−0.08 to 1.44	.50

*Values are estimates.

†*P* value is for the coefficient using the Wald Z statistic.

Discussion

In summary, our descriptive results show that peer reviews have not been integrated into the development process for 63% (95% CI, 54-72) of the projects, while testing has been integrated into the development of 82% (95% CI, 75-89) of the projects. Approximately 74% (95% CI, 67-81) of developers did not have a background in computing, 80% (95% CI, 74-87) were paid for their contributions to the project, and 52% (95% CI, 43-60) had PhDs. A multivariate logistic regression model to predict the implementation of peer reviews was not significant (likelihood ratio test = 16.86, 9 df, $P = .051$) and neither was a model to predict the implementation of testing (likelihood ratio test = 3.34, 9 df, $P = .95$).

Developer Background

The level of experience of the biomedical OS developers is consistent with the experience of developers of general (nonbiomedical) OS projects. Previous surveys found that developers have, on average, 11.86 years of programming experience [81]. However, compared to nonbiomedical OS projects, the biomedical developers tended to have less software engineering experience, but more advanced degrees. Previous surveys of general OS developers found that software engineers and programmers made up 43% of OS developers [87], 45.4% of OS developers were programmers [88], almost 80% of them worked in the IT sector [87], 58% were directly involved in the IT industry [81], 45% worked as professional programmers [81], and 51% had formal university level training in computer science and IT [81]. In addition, surveys of general OS

developers found that a quarter of respondents had only high school or grammar school education [79,80], only 5% had PhDs, and 12% had masters degrees [87].

The lack of formal training and education in software development practices by the biomedical OS developers raises questions about the quality of the software being developed. However, our logistic regression models control for project characteristics and they did not find a relationship between programmer background/experience and the extent of implementation of peer reviews and testing. Therefore, it does not seem that the lack of formal training and education in software engineering has affected the implementation of good QA practices in these biomedical OS projects.

The percentage of developers financially supported to develop the biomedical OS applications is much larger than that seen in nonbiomedical OS projects. For example, other surveys of general OS developers found that 16% [79,80], 20% [87,89], 30% [88], and 40% [81] of developers are paid for their OS contributions. In our survey, we found around 80% of developers were being paid for their contributions. It is not clear from our results whether there were any commercial interests financing such development work.

Implementation of Peer Reviews

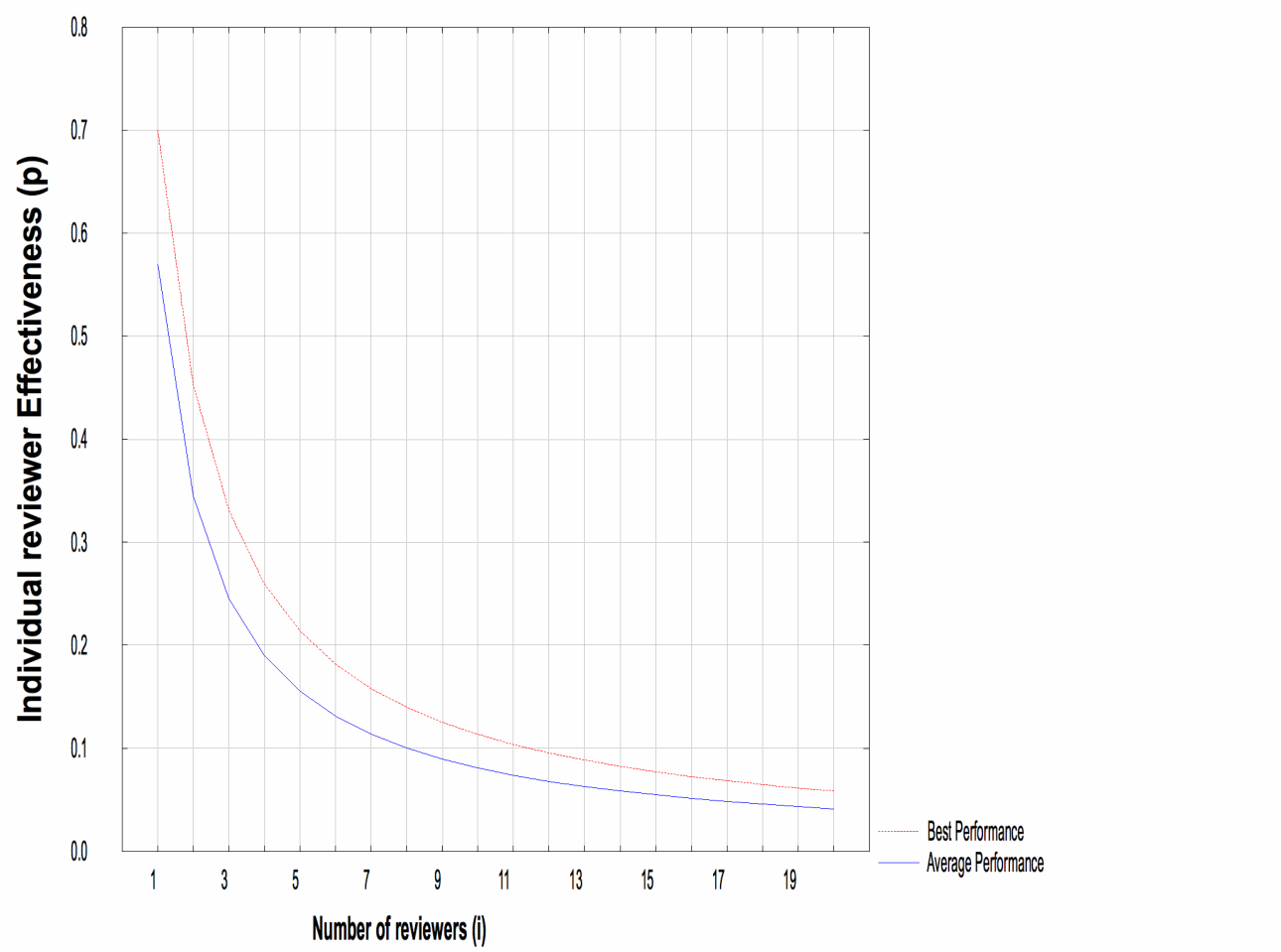
Even though extensive peer reviews are often claimed to be one of the main advantages of the OS development paradigm, our results indicate that this practice is not prevalent, with the majority of projects not undergoing peer review on a consistent basis, and two fifths never doing so. This finding is somewhat consistent with peer review of general (nonbiomedical) OS projects, which was summarized earlier in the paper.

For projects that perform peer reviews, how well is the peer review implemented? A key performance measure for a peer review is the proportion of defects that it finds in the code (this is known as the effectiveness of the peer review). There are a number of factors that will have an influence on that: the number of reviewers, the capability of the reviewers (often measured by the proportion of defects that an individual reviewer can find), and the reading technique that is used.

We found in our survey that a maximum of three reviewers review a piece of code when it is peer reviewed. If we assume that reviewers are independent, which is a reasonable assumption in a distributed development project with little to no face-to-face interaction, a basic model of the probability of finding a defect is given by $1 - (1 - p)^i$ where p is the probability that an individual reviewer will find a defect (assuming that all reviewers are equally capable) and i is the number of reviewers. A previous review of the literature [90] determined that the average probability of finding a defect through code peer review was 0.57 and the maximum (or best-in-class) was 0.7. These numbers came from industrial, non-OS projects.

Figure 1 shows the theoretical relationship based on the above model between the number of reviewers (x-axis) and individual reviewer effectiveness (y-axis) if we fix the overall peer review effectiveness at 0.57 and at 0.7. We can see that if a team of at most three programmers reviews a code snippet, then they would each have to have a p of at least 0.24 to achieve average performance, and a p of 0.33 to achieve maximum (best-in-class) performance. These would be the minimal reviewer capabilities for a team of three to achieve the average and maximum peer review effectiveness reported for non-OS projects, respectively. Are these minimal capabilities plausible, that is, is it plausible that the biomedical OS reviewers achieve defect detection effectiveness levels that high?

Figure 1. Relationship between the number of reviewers and individual effectiveness when the team effectiveness is fixed at 0.57 (average performance) and 0.7 (best performance)



The literature on individuals' effectiveness in code reviews can be examined to answer this question. Wohlin et al [91] created virtual review teams using the data sets collected from the literature, and they found that the effectiveness of individual reviewers had a median value close to 0.25. In another study, Runeson and Wohlin [92] reported that the defect detection rates observed during their experiments involving students and professionals had a median value of 0.31. Land et al [93] reported that, in their experiment, an individual detected an average of 5.51 out of 33 defects (effectiveness of approximately 0.17). Porter et al [94] reported the true positive ratios as a measure of effectiveness instead of the detection rates. On average, 13% of the reported issues turned out to be true defects. Dunsmore et al [50] conducted a code inspection experiment for Java programs in which they reported both detection ratio and false positives. The detection ratio for checklist-based reading was 52.14%, and the false positive rate was 24.50%. In this experiment, the subjects reviewed 200 lines of Java code with which they had familiarity from previous exercises, and they were provided with class diagrams and natural language specifications of all systems. In Land et al's experiment, the subjects were provided with flowcharts, pseudo-code, and other code overview documents. In realistic OS development environments, such aids are less likely to be available. Therefore, it will be more difficult for the OS developers to reach the effectiveness levels mentioned in these studies.

Even if it is assumed that it is plausible for a biomedical OS peer reviewer to achieve average effectiveness rates as high as those achieved by non-OS peer reviewers, it would be less likely that OS projects would achieve the maximum or best-in-class effectiveness rates seen in non-OS projects. In addition, if the number of reviewers dips below three, then it is not likely that the effectiveness of the peer review would match the average performance of the non-OS projects.

That very few of the projects use checklists during the code reading also indicates that the effectiveness of these peer reviews will be relatively low.

Implementation of Testing

How do testing activities in biomedical OS projects compare to testing in general OS projects? Zhao and Elbaum [53] noted that 58% of the generic OS projects spent more than 20% of their development time in testing. We found that 51% of the biomedical OS projects spent more than 20% of their time in testing. In generic OS projects, almost 30% had below 30% code coverage [53]. For 35.85% of the biomedical OS projects, the coverage was either not known or it was estimated as below 20%. The use of a regression test suite in biomedical OS projects was almost 58% compared to 3% in very large generic OS projects. In summary, testing activities in biomedical OS projects showed similarity to those in the generic OS projects.

However, compared to the closed-source projects [34], there is still room for improvement.

Summary

The major QA risk item in biomedical OS software projects is that code peer reviews are not systematically performed. Peer reviews are an important mechanism to find bugs in software, and they are complementary to testing.

Applications that collect and manage patient data can lead to negative financial and patient safety outcomes if they have subtle defects in them. OS developers therefore need to focus more attention on integrating peer review into their development practices to ensure that good software engineering practices are systematically employed. Acquirers of biomedical OS applications need to ensure that some form of peer review is being consistently practiced in the software projects producing and maintaining the applications that they deploy.

Limitations

We highlight two limitations in this study. The first is the low response rate, although the response rate that we obtained is consistent with other studies in the same domain. In addition, we found no evidence of nonresponse bias. Second, there are many OS projects that were not included in our sample, and it is plausible that the ones we focused on tended to be the smaller ones.

Conclusions

The OS software development paradigm has been suggested as a new and more effective way to develop high-quality software. In many health-related settings, ranging from care to research, it is important to ensure that software failures are minimized.

In this paper we performed a survey of biomedical OS software developers to understand their QA practices. Our results indicate that the major risk item in biomedical OS projects, from a software quality perspective, is the (lack of) implementation of peer reviews. Furthermore, when they are implemented, their performance is below what would be considered best practice. We also found that most of the developers did not have computer science or computer engineering training or education that could provide them with software engineering background. On the other hand, we found no evidence linking the lack of computer science or engineering background with the extent of implementation of peer reviews and testing, indicating that such background variables do not have an impact.

These results highlight some risk from transitioning biomedical OS applications into environments where they may have an impact on patient safety. For this transition to occur, it is important that better peer review practices be put in place. To the extent possible, developers of biomedical OS software should rely on Food and Drug Administration regulations and guidelines, such as 21 CFR Part 11, as well as professional society publications [95] documenting what are considered to be best software engineering practices.

Acknowledgments

This work was partially funded by a grant from the Natural Sciences and Engineering Research Council of Canada and by a research grant from Microsoft Corporation. We would like to thank our colleagues in the Georgetown Medical Center and the National Cancer Institute for participating in the pilot survey and providing feedback, and Bryan Kirschner for feedback on an earlier version of this paper.

Conflicts of Interest

Khaled El Emam has developed and distributed OS software under the GNU GPL v2 license. He is also a co-founder and has financial interests in TrialStat Corporation, which is a commercial software company in the biomedical domain.

Multimedia Appendix 1

Survey form [PDF File, 81 KB - jmir_v9i2e8_app1.pdf]

Multimedia Appendix 2

List of OS projects contacted for survey [PDF File, 212 KB - jmir_v9i2e8_app2.pdf]

Multimedia Appendix 3

CHERRIES [84] summary for the Web survey [PDF File, 105 KB - jmir_v9i2e8_app3.pdf]

References

1. Prada G, Roberts G, Vail S, Anderson M, Down E, Fooks G, et al. Understanding Health Care Cost Drivers and Escalators. Ottawa, ON: The Conference Board of Canada; 2004. URL: http://www.health.gov.ab.ca/resources/publications/conference_board.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5NyNE0fv9]

2. Prada G, Grimes K, McCleery A, Nguyen D, Pomey MP, Reed V, et al. Challenging Health Care System Sustainability: Understanding Health System Performance of Leading Countries. Ottawa, ON: The Conference Board of Canada; 2004. URL: http://www.health.gov.ab.ca/resources/publications/Conference_Board2.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5Nt0athGf]
3. Irving R. Report on IT in Canadian Hospitals. Canadian Healthcare Technology 2003.
4. Raymond B, Dold C. Clinical Information Systems: Achieving the Vision. Oakland, CA: Kaiser Permanente Institute for Health Policy; 2002. URL: http://www.kpihp.org/publications/docs/clinical_information.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5Nt0d8r77]
5. Bates D, Ebell M, Gotlieb E, Zapp J, Mullins H. A proposal for electronic medical records in US primary care. J Am Med Inform Assoc 2003;10(1):1-10. [doi: 10.1197/jamia.M1097]
6. Goulde M, Brown E. Open Source Software: A Primer for Health Care Leaders. Oakland, CA: California HealthCare Foundation; 2006. URL: http://www.chcf.org/documents/ihealth/OpenSourcePrimer.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5Nt0g8WP4]
7. Kantor GS, Wilson WD, Midgley A. Open-source software and the primary care EMR. J Am Med Inform Assoc 2003;10(6):616 [FREE Full text] [Medline: 12925540] [doi: 10.1197/jamia.M1403]
8. Giera J. The Costs and Risks of Open Source. Cambridge, MA: Forrester Research Inc; 2004.
9. Hunt F, Probert D, Barratt S. Adopting new technology: the case of open source software at Marconi. The 12th International Conference on Management of Technology (IAMOT May 12-15 2003) Nancy, France 2003.
10. Dal MJ. Open Source Software in Canada. A Collaborative Fact Finding Study. e-cology Corporation. 2003. URL: http://www.e-cology.ca/canfloss/report/CANfloss_Report.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5Nv7MQJQa]
11. Dedrick J, West J. An exploratory study into open source platform adoption. Proceedings of the 37th Hawaii International Conference on System Sciences 2004:10.
12. DiDio L. Linux, Unix and Windows TCO Comparison, Part 1. Boston, MA: The Yankee Group; 2004.
13. Valdes I, Kibbe DC, Tolleson G, Kunik ME, Petersen LA. Barriers to proliferation of electronic medical records. Inform Prim Care 2004;12(1):3-9. [Medline: 15140347]
14. Kolata G. In unexpected Medicare benefit, US will offer doctors free electronic records system. New York Times Jul 21 2005:14.
15. Goldstein D, Ponkshe S, Maduro R. Analysis of Open Source Software (OSS) and EHRs: Profile of Increasing Use of OSS in the Federal Government and Healthcare. Atlantis, FL: Medical Alliances Inc; 2004. URL: http://www.medicalalliances.com/downloads/files/Open_Source_Software-Government_and_Healthcare_White_Paper-Medical_Alliances_2.doc [accessed 2007 Apr 16] [WebCite Cache ID 5Nt0If4XE]
16. Elsner C, Egbring M, Kottkamp H, Berger T, Zoller S, Hinricks G. Open source or commercial products for electronic data capture in clinical trials? A scorecard comparison. Comput Cardiol 2003;30:371-373. [doi: 10.1109/CIC.2003.1291169]
17. Erickson BJ, Langer S, Nagy P. The role of open-source software in innovation and standardization in radiology. J Am Coll Radiol 2005 Nov;2(11):927-931. [Medline: 17411967] [doi: 10.1016/j.jacr.2005.05.004]
18. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al; Yang JYH. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5(10):R80.
19. Stajich J. The Bioperl Project: A Look Ahead. Talk presented at: Bioinformatics Open Source Conference; August 1-2, Edmonton, Canada. 2002. URL: http://www.open-bio.org/bosc2002/slides/2002-08-01-Stajich-Bioperl.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5Nv7MQJQa]
20. Pocock M. BioJava Toolkit Progress. Talk presented at: Bioinformatics Open Source Conference; August 1-2,. 2002. URL: http://www.open-bio.org/bosc2002/talks.html#biojava [accessed 2007 Apr 16] [WebCite Cache ID 5Nsk5oct4]
21. McCormick J, Gage D. Cincinnati Children's Hospital: Shots in the Dark. Baseline. 2004 Aug 1. URL: http://www.baselinemag.com/article2/0,1540,1655082,00.asp [accessed 2007 Apr 16]
22. Gage D, McCormick J. Why Software Quality Matters: 'We Did Nothing Wrong'. Baseline. 2004 Mar 4. URL: http://www.baselinemag.com/article2/0,1397,1544403,00.asp [accessed 2007 Apr 16]
23. Basili V, Belady L, Boehm B, Brooks F, Browne J, DeMillo R, et al. Final report: NSF workshop on a software research program for the 21st century. Software Eng Notes 1999;24(3):37-45.
24. ; Report of the Inquiry into the London Ambulance Service. UK: South West Thames Regional Health Authority; 1993. URL: http://www.cs.ucl.ac.uk/staff/A.Finkelstein/las/lascase0.9.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5NskKydHy]
25. van Heusden P. Applying Software Validation Techniques to Bioperl. Talk presented at: Bioinformatics Open Source Conference; July 29-30, 2004;Glasgow, UK. URL [FREE Full text]
26. Konrad M, Shrum S; Chrissis M-B. CMMI: Guidelines for Process Integration and Product Improvement. Addison-Wesley; 2003.
27. Raymond ES. The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly; 1999.
28. Greiner S, Boskovic B, Brest J, Zumer V. Security issues in information systems based on open source technologies. EUROCON 2003:12-15.

29. El EK. Software Inspection Best Practices. In: Agile Project Management Advisory Service, Executive Report. Volume 2. Number 9. Arlington, MA: Cutter Consortium; 2001.
30. Bagchi S, Madeira H. Open source software - a recipe for vulnerable software, or the only way to keep the bugs and the bad guys out? Panel position statement presented at: 14th International Symposium on Software Reliability Engineering (ISSRE) 2003; 1234. 567:347-352.
31. Jorgensen N. Putting it all in the trunk: incremental software development in the FreeBSD open source project. Inform Syst J 2001;11(4):321-336. [doi: 10.1046/j.1365-2575.2001.00113.x]
32. Stark J. Peer reviews as a quality management technique in open-source software development projects. European Conference on Software Quality 2002:340-350.
33. Maccormack A, Kemerer C, Cusumano M, Crandall B. Trade-offs between productivity and quality in selecting software development practices. IEEE Software 2003;(5):78-84. [doi: 10.1109/MS.2003.1231158]
34. Cusumano M, MacCormack A, Kemerer C, Randall B. Software development worldwide: the state of the practice. IEEE Software 2003; 1234. 567:28-34.
35. Zhao L, Elbaum S. A survey on quality related activities in open source. Software Eng Notes 2000;25(3):54-57. [doi: 10.1145/505863.505878]
36. Porter A, Siy H, Toman C, Votta L. An experiment to assess the cost-benefits of code inspections in large scale software development. IEEE T Software Eng 1997;23(6):329-346. [doi: 10.1109/32.601071]
37. STD I. IEEE Standard for Software Reviews and Audits. ANSI/IEEE STD 1028-1988 1989:.
38. Ackerman AF, Buchwald LS, Lewski FH. Software inspections: an effective verification process. IEEE Software 1989;6(3):31-36. [doi: 10.1109/52.28121]
39. Lau L, Jeffery R, Sauer C. Some Empirical Support for Software Development Technical Reviews. Sydney, Australia: University of New South Wales; 1996.
40. Grady RB. Practical Software Metrics for Project Management and Process Improvement. Englewood Cliffs, NJ: Prentice-Hall; 1992.
41. Laitenberger O, Debaud JM. An encompassing life-cycle centric survey of software inspection. J Syst Software 2000;50(1):5-31. [doi: 10.1016/S0164-1212(99)00073-4]
42. Fagan ME. Design and code inspections to reduce errors in program development. IBM Syst J 1976;15(3):182-211.
43. Weller EF. Lessons from three years of inspection data. IEEE Software 1993;10(5):38-45. [doi: 10.1109/52.232397]
44. Madachy R, Little L, Fan S. Analysis of a successful inspection program. Proceedings of the 18th Annual NASA Software Engineering Laboratory Workshop 1993:176-188.
45. Gilb T, Graham D. Software Inspection. New York: Addison-Wesley; 1993.
46. Strauss S, Ebenau R. Software Inspection Process. New York: McGraw Hill; 1994.
47. Kelly J, Sheriff J, Hops J. An analysis of defect densities found during software inspections. J Syst Software 1992;17(2):111-117. [doi: 10.1016/0164-1212(92)90089-3]
48. Johnson P. Reengineering inspection. Comm ACM 1998;41(2):49-52. [doi: 10.1145/269012.269020]
49. Eick S, Loader C, Long M, Votta L, Weil SV. Estimating software fault content before coding. Proceedings of the 14th International Conference on Software Engineering 1992:59-65.
50. Dunsmore A, Roper M, Wood M. The development and evaluation of three diverse techniques for object-oriented code inspection. IEEE T Software Eng 2003;29(8):677-686. [doi: 10.1109/TSE.2003.1223643]
51. Halloran T, Scherlis WL. High quality and open source software practices. Position paper presented at: The Second Workshop on Open Source Software Engineering, International Conference on Software Engineering; May 19-25,. 2002. URL: http://opensource.ucc.ie/icse2002/HalloranScherlis.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5NskgWv6b]
52. Gacek C, Arief B. The many meanings of open source. IEEE Software 2004;21(1):34-40. [doi: 10.1109/MS.2004.1259206]
53. Zhao L, Elbaum S. Quality assurance under the open source development model. J Syst Software 2003;66(1):65-75.
54. Mockus A, Fielding R, Herbsleb J. Two case studies of open source software development: Apache and Mozilla. ACM T Softw Eng Meth 2002;11(3):309-346. [doi: 10.1145/567793.567795]
55. Sharma S, Sugumaran V, Rajagoplan B. A framework for creating hybrid open source software communities. Inform Syst J 2002;12(1):7-25. [doi: 10.1046/j.1365-2575.2002.00116.x]
56. Erenkrantz J. Release management within open source projects. Proceedings of the Third Workshop on Open Source Software Engineering, Portland, OR 2003.
57. Thomas C. Improving verification, validation, and test of the Linux kernel: the Linux stabilization project. In: Proceedings of the Third Workshop on Open Source Software Engineering. Portland, OR; 2003.
58. Iyer M. Analysis of Linux test project's kernel code coverage. Austin, TX: IBM Corporation; 2002. URL: http://ltp.sourceforge.net/documentation/technical_papers/kernel_coverage.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5NvKVXOxu]
59. Myers G. A controlled experiment in program testing and code walkthroughs/inspections. Comm ACM 1978;21(9):760-768. [doi: 10.1145/359588.359602]
60. Basili V, Selby R. Comparing the effectiveness of software testing strategies. IEEE T Software Eng 1987;13(12):1278-1296.

61. Kamsties E, Lott C. An empirical evaluation of three defect-detection techniques. Proceedings of the 5th European Software Engineering Conference, Sitges, Spain, September 25-28, 1995.
62. Wood M, Roper M, Brooks A, Miller J. Comparing and combining software defect-detection techniques. Proceedings of the 6th European Conference on Software Engineering 1997:262-277.
63. Jalote P, Haragopal M. Overcoming the NAH Syndrome for Inspection Deployment. Proceedings of the International Conference on Software Engineering 1998:371-378. [doi: 10.1109/ICSE.1998.671390]
64. Hissam S, Weinstock C, Plakosh D, Asundi J. Perspectives on Open Source Software. CMU/SEI-2001-TR-019. Pittsburg, PA: Carnegie Mellon Software Engineering Institute; 2001. URL: http://www.sei.cmu.edu/pub/documents/01.reports/pdf/01tr019.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5NslnRroS]
65. Mockus A, Fielding RT, Herbsleb J. The Apache server.. Proceedings of the 22nd International Conference on Software Engineering A case study of open source software development; 2000:263-272.
66. Miller B, Fredriksen L, So B. An Empirical Study of the Reliability of Unix Utilities. Comm ACM 1990;33(12):32-44. [doi: 10.1145/96267.96279]
67. Capiluppi A, Lago P, Morisio M. Characteristics of open source projects. Proceedings of the Seventh European Conference on Software Maintenance and Engineering; 2003;:317. [doi: 10.1109/CSMR.2003.1192440]
68. Hunt F, Johnson P. On the Pareto Distribution of Sourceforge Projects. Cambridge, UK: Cambridge University; 2003. URL: http://www.ifm.eng.cam.ac.uk/people/fhh10/Sourceforge/Sourceforge%20paper.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5NslBEpHN]
69. Dinh-Trong T, Bieman J. The FreeBSD project: a replication case study of open source development. IEEE T Software Eng 2005;31(6):481-494. [doi: 10.1109/TSE.2005.73]
70. Stark JE. Peer Reviews in Open-Source Software Development. Australia: Griffith University; 2001. URL: http://ecommerce.cit.gu.edu.au/ict/docs/theses/JStark_Dissertation_OSS.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5NslHVMTe]
71. SourceForge Home page. sourceforge.net. URL: http://sourceforge.net/ [accessed 2007 Apr 16] [WebCite Cache ID 5HrhxjQYC]
72. FreshMeat Home page. freshmeat.net. URL: http://freshmeat.net/ [accessed 2007 Apr 16] [WebCite Cache ID 5HrjAMLJ4]
73. BioMed Central Home page. biomedcentral.com. URL: http://www.biomedcentral.com/ [accessed 2007 Apr 16] [WebCite Cache ID 5Hri4z1L8]
74. Koru AG, Tian J. Defect Handling in Medium and Large Open Source Projects. IEEE Software 2004;21(4):54-61. [doi: 10.1109/MS.2004.12]
75. Kreft I, Yoon B. Are multilevel techniques necessary? An attempt at demystification. Annual Meeting of the American Educational Research Association; April 4-8,. New Orleans, LA; 1994. URL: http://eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/26/c1/54.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5NslPObfA]
76. Maas C, Hox J. Sufficient Sample Sizes for Multilevel Modeling. Methodology 2005;1(3):86-92. [doi: 10.1027/1614-2241.1.3.85]
77. Hosmer D, Lemeshow S. Applied Logistic Regression. New York: John Wiley & Sons; 1989.
78. Schonlau M, Fricker RD, Elliott MN. Conducting Research Surveys via E-mail and the Web. Santa Monica, CA: RAND Corporation; 2002.
79. Hars A, Ou S. Working for free? Motivations for participating in open source projects. Int J Electron Comm 2002;6(3):25-39.
80. Hars A, Ou S. Working for free? Motivations of participating in open source projects. Proceedings of the 34th Hawaii International Conference on System Sciences 2001;7:7014.
81. Lakhani K, Wolf R. Why hackers do what they do: understanding motivation and effort in free/open source software projects. In: Feller J, Fitzgerald B, Hissam S, Lakhani K, editors. Perspectives on Free and Open Source Software. Cambridge, MA; The MIT Press; 2005.
82. Lindner JR, Murphy TH, Briers GE. Handling nonresponse in social science research. J Agr Educ 2001;42(4):43-53.
83. Siegel S, Castellan J. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw Hill; 1988.
84. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [Medline: 15471760] [doi: 10.2196/jmir.6.3.e34]
85. Serlin R, Levin J. Teaching how to derive directly interpretable coding schemes for multiple regression analysis. J Educ Stat 1985;10(3):223-238. [doi: 10.2307/1164794]
86. Wendorf C. Primer on multiple regression coding: Common forms and the additional case of repeated contrasts. Understand Stat 2004;3(1):47-57. [doi: 10.1207/s15328031us0301_3]
87. Robles G, Scheider H, Tretkowski I, Weber N. Who is doing it? A research on Libre Software developers. Berlin, Germany: Technical University of Berlin; 2001. URL: http://widi.berlios.de/paper/study.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5NsmG1QA7]
88. Lakhani K, Wolf B, Bates J, DiBona C. The Boston Consulting Group hacker survey. O'Reilly Open Source Conference; July 22-26,. 2002. URL: http://www.ostg.com/bcg/BCGHACKERSURVEY.pdf [accessed 2007 Apr 16] [WebCite Cache ID 5NslYu4Ht]
89. Hertel G, Niedner S, Herrmann S. Motivation of software developers in open source projects: an Internet-based survey of contributors to the Linux kernel. Res Pol 2003;32(7):1159-1177. [doi: 10.1016/S0048-7333(03)00047-7]

90. Briand L, El Emam K, Laitenberger O, Fussbroich T. Using simulation to build inspection efficiency benchmarks for development projects. Proceedings of the 20th International Conference on Software Engineering 1998:340-349. [doi: 10.1109/ICSE.1998.671387]
91. Wohlin C, Aurum A, Petersson H, Shull F, Ciolkowski M. Software inspection benchmarking - a qualitative and quantitative comparative opportunity. METRICS '02 Proceedings of the 8th International Symposium on Software Metrics 2002:118-127. [doi: 10.1109/METRIC.2002.1011331]
92. Runeson P, Wohlin C. An experimental evaluation of an experience-based capture-recapture method in software code inspections. Empir Softw Eng 1998;3(4):381-406. [doi: 10.1023/A:1009728205264]
93. Sauer C, Jeffery R; Land LPW. Validating the defect detection performance advantage of group designs for software reviews: report of a laboratory experiment using program code. In: ESEC '97/FSE-5: Proceedings of the 6th European conference held jointly with the 5th ACM SIGSOFT international symposium on Foundations of software engineering 1997:294-309-294-309.
94. Porter A, Siy H, Toman C, Votta L. An experiment to assess the cost-benefits of code inspections in large scale software development. Proceedings of the 3rd ACM SIGSOFT Symposium on Foundations of Software Engineering 1995;:92-103.
95. ; The Good Automated Manufacturing Practice (GAMP) Guide for Validation of Automated Systems in Pharmaceutical Manufacture. Tampa, FL: International Society for Pharmaceutical Engineering; 2002.

submitted 22.08.06; peer-reviewed by P Murray, R Pietrobon, C McDonald; comments to author 05.09.06; revised version received 09.04.07; accepted 11.04.07; published 07.05.07.

Please cite as:
Koru G, El Emam K, Neisa A, Umarji M
A Survey of Quality Assurance Practices in Biomedical Open Source Software Projects
J Med Internet Res 2007;9(2):e8
URL: http://www.jmir.org/2007/2/e8/
doi:10.2196/jmir.9.2.e8
PMID:17513286

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