Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 13 (2011), Issue 4 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Editorials

Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact (e123)
Gunther Eysenbach. . . . 5

CONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions (e126)
Gunther Eysenbach, CONSORT-EHEALTH Group. . . . 25

Guest Editorials

WikiBuild: A New Application to Support Patient and Health Care Professional Involvement in the Development of Patient Support Tools (e114)
Patrick Archambault. . . . 35

eHealth Literacy 2.0: Problems and Opportunities With an Evolving Concept (e125)
Cameron Norman. . . . 39

Viewpoints

A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies (e111)
Julia van Gemert-Pijnen, Nicol Nijland, Maarten van Limburg, Hans Ossebaard, Saskia Kelders, Gunther Eysenbach, Erwin Seydel. . . . 43

Why Business Modeling is Crucial in the Development of eHealth Technologies (e124)
Maarten van Limburg, Julia van Gemert-Pijnen, Nicol Nijland, Hans Ossebaard, Ron Hendrix, Erwin Seydel. . . . 62

Ethical Principles for Physician Rating Sites (e113)
Daniel Strech. . . . 415

XSL•FO
RenderX

Original Papers

Weight Change in a Commercial Web-Based Weight Loss Program and its Association With Website Use: Cohort Study (e83)
Melinda Neve, Philip Morgan, Clare Collins. . . . 73

Effect of Tailoring in an Internet-Based Intervention for Smoking Cessation: Randomized Controlled Trial (e121)
Silje Wangberg, Olav Nilsen, Konstantinos Antypas, Inger Gram. . . . 83

Development of an Interactive, Web-Delivered System to Increase Provider–Patient Engagement in Smoking Cessation (e87)
Rajani Sadasivam, Kathryn Delaughter, Katie Crenshaw, Heather Sobko, Jessica Williams, Heather Coley, Midge Ray, Daniel Ford, Jeroan Allison, Thomas Houston. . . . 95

Health Promotion in the Workplace: Assessing Stress and Lifestyle With an Intranet Tool (e88)
Daniela Lucini, Nadia Solaro, Alessandro Lesma, Veronique Gillet, Massimo Pagani. . . . 107

Improvement of Physical Activity by a Kiosk-based Electronic Screening and Brief Intervention in Routine Primary Health Care: Patient-Initiated Versus Staff-Referred (e99)
Matti Leijon, Daniel Arvidsson, Per Nilsen, Diana Stark Ekman, Siw Carlfjord, Agneta Andersson, Anne Johansson, Preben Bendtsen. . . . 128

Web-Based Risk Communication and Planning in an Obese Population: Exploratory Study (e100)
Anastasia Soureti, Peter Murray, Mark Cobain, Willem van Mechelen, Robert Hurling. . . . 139

Evaluation of a Web-Based Intervention to Promote Hand Hygiene: Exploratory Randomized Controlled Trial (e107)
Lucy Yardley, Sascha Miller, Wolff Schlotz, Paul Little. . . . 168

Exploratory Study of Web-Based Planning and Mobile Text Reminders in an Overweight Population (e118)
Anastasia Soureti, Peter Murray, Mark Cobain, Mai Chinapaw, Willem van Mechelen, Robert Hurling. . . . 182

Tweets, Apps, and Pods: Results of the 6-Month Mobile Pounds Off Digitally (Mobile POD) Randomized Weight-Loss Intervention Among Adults (e120)
Gabrielle Turner-McGrievy, Deborah Tate. . . . 193

Effectiveness of Web-Based Versus Face-To-Face Delivery of Education in Prescription of Falls-Prevention Exercise to Health Professionals: Randomized Trial (e116)
Stephen Maloney, Romi Haas, Jennifer Keating, Elizabeth Molloy, Brian Jolly, Jane Sims, Prue Morgan, Terry Haines. . . . 207

Hospital-Based Nurses' Perceptions of the Adoption of Web 2.0 Tools for Knowledge Sharing, Learning, Social Interaction and the Production of Collective Intelligence (e92)
Adela Lau. . . . 217

Online Social Networks and Smoking Cessation: A Scientific Research Agenda (e119)
Nathan Cobb, Amanda Graham, M. Byron, Raymond Niaura, David Abrams, Workshop Participants. . . . 229

Anxiety Online—A Virtual Clinic: Preliminary Outcomes Following Completion of Five Fully Automated Treatment Programs for Anxiety Disorders and Symptoms (e89)
Britt Klein, Denny Meyer, David Austin, Michael Kyrios. . . . 245

Changes in Depressive Symptoms, Social Support, and Loneliness Over 1 Year After a Minimum 3-Month Videoconference Program for Older Nursing Home Residents (e93)
Hsiu-Hsin Tsai, Yun-Fang Tsai. . . . 265

XSL•FO
RenderX

Impact of Length or Relevance of Questionnaires on Attrition in Online Trials: Randomized Controlled Trial (e96)
Jim McCambridge, Eleftheria Kalaitzaki, Ian White, Zarnie Khadjesari, Elizabeth Murray, Stuart Linke, Simon Thompson, Christine Godfrey, Paul Wallace. . . . 277

Methodological Issues in Internet-Mediated Research: A Randomized Comparison of Internet Versus Mailed Questionnaires (e109)
Lisa Whitehead. . . . 286

Internet Use for Health-Related Information via Personal Computers and Cell Phones in Japan: A Cross-Sectional Population-Based Survey (e110)
Yoshimitsu Takahashi, Tomoko Ohura, Tatsuro Ishizaki, Shigeru Okamoto, Kenji Miki, Mariko Naito, Rie Akamatsu, Hiroki Sugimori, Nobuo Yoshiike, Koichi Miyaki, Takuro Shimbo, Takeo Nakayama. . . . 292

Improving Access to Information and Support for Patients With Less Common Cancers: Hematologic Cancer Patients' Views About Web-Based Approaches (e112)
Christine Paul, Mariko Carey, Alix Hall, Marita Lynagh, Robert Sanson-Fisher, Frans Henskens. . . . 306

Using Natural Language Processing to Enable In-depth Analysis of Clinical Messages Posted to an Internet Mailing List: A Feasibility Study (e98)
Tanja Bekhuis, Marcos Kreinacke, Heiko Spallek, Mei Song, Jean O'Donnell. . . . 316

WikiBuild: A New Online Collaboration Process For Multistakeholder Tool Development and Consensus Building (e108)
Samir Gupta, Flora Wan, David Newton, Onil Bhattacharyya, Mark Chignell, Sharon Straus. . . . 328

Beyond Readability: Investigating Coherence of Clinical Text for Consumers (e104)
Catherine Smith, Scott Hetzel, Prudence Dalrymple, Alla Keselman. . . . 345

Analysis of the Definition and Utility of Personal Health Records Using Q Methodology (e105)
Jeongeun Kim, David Bates. . . . 362

Determinants of Engagement in Face-to-Face and Online Patient Support Groups (e106)
Cornelia Van Uden-Kraan, Constance Drossaert, Erik Taal, Willem Smit, Hein Bernelot Moens, Mart Van de Laar. . . . 371

Benefits of Peer Support in Online Japanese Breast Cancer Communities: Differences Between Lurkers and Posters (e122)
Yoko Setoyama, Yoshihiko Yamazaki, Kazuhiro Namayama. . . . 381

Website Quality, Expectation, Confirmation, and End User Satisfaction: The Knowledge-Intensive Website of the Korean National Cancer Information Center (e81)
Chulmo Koo, Yulia Wati, Keeho Park, Min Lim. . . . 394

Communications Between Volunteers and Health Researchers during Recruitment and Informed Consent: Qualitative Content Analysis of Email Interactions (e84)
Anne Townsend, Zubin Amarsi, Catherine Backman, Susan Cox, Linda Li. . . . 408

Digital Dashboard Design Using Multiple Data Streams for Disease Surveillance With Influenza Surveillance as an Example (e85)
Calvin Cheng, Dennis Ip, Benjamin Cowling, Lai Ho, Gabriel Leung, Eric Lau. . . . 422

Attrition in Web-Based Treatment for Problem Drinkers (e117)
Marloes Postel, Hein de Haan, Elke ter Huurne, Job van der Palen, Eni Becker, Cor de Jong. . . . 431

XSL•FO
RenderX

Using Virtual Reality to Provide Health Care Information to People With Intellectual Disabilities: Acceptability, Usability, and Potential Utility (e91)
Valerie Hall, Suzanne Conboy-Hill, Dave Taylor. . . . 445

Development and Validation of Filters for the Retrieval of Studies of Clinical Examination From Medline (e82)
Nader Shaikh, Robert Badgett, Mina Pi, Nancy Wilczynski, K. McKibbon, Andrea Ketchum, R. Haynes. . . . 458

Do Family Physicians Retrieve Synopses of Clinical Research Previously Read as Email Alerts? (e101)
Roland Grad, Pierre Pluye, Janique Johnson-Lafleur, Vera Granikov, Michael Shulha, Gillian Bartlett, Bernard Marlow. . . . 469

Effects of an eHealth Literacy Intervention for Older Adults (e90)
Bo Xie. . . . 480

Does the eHealth Literacy Scale (eHEALS) Measure What it Intends to Measure? Validation of a Dutch Version of the eHEALS in Two Adult Populations (e86)
Rosalie van der Vaart, Alexander van Deursen, Constance Drossaert, Erik Taal, Jan van Dijk, Mart van de Laar. . . . 499

A Framework for Characterizing eHealth Literacy Demands and Barriers (e94)
Connie Chan, David Kaufman. . . . 510

Analysis of 4999 Online Physician Ratings Indicates That Most Patients Give Physicians a Favorable Rating (e95)
Bassam Kadry, Larry Chu, Bayan Kadry, Danya Gammas, Alex Macario. . . . 538

Physician Response Time When Communicating With Patients Over the Internet (e79)
Per Kummervold, Jan-Are Johnsen. . . . 547

Public Access and Use of Health Research: An Exploratory Study of the National Institutes of Health (NIH) Public Access Policy Using Interviews and Surveys of Health Personnel (e97)
Jamie O'Keeffe, John Willinsky, Lauren Maggio. . . . 553

A Study of Innovative Features in Scholarly Open Access Journals (e115)
Bo-Christer Björk. . . . 562

Reviews

Effects of eHealth Interventions on Medication Adherence: A Systematic Review of the Literature (e103)
Annemiek Linn, Marcia Vervloet, Liset van Dijk, Edith Smit, Julia Van Weert. . . . 152

eHealth Literacy Among College Students: A Systematic Review With Implications for eHealth Education (e102)
Michael Stellefson, Bruce Hanik, Beth Chaney, Don Chaney, Bethany Tennant, Enmanuel Chavarria. . . . 526

XSL•FO
RenderX

Editorial

Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact

Gunther Eysenbach[1,2,3], MD, MPH, FACMI

[1]University Health Network, Centre for Global eHealth Innovation & Techna Institute, Toronto, ON, Canada

[2]Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

[3]JMIR Publications Inc., Toronto, ON, Canada

Corresponding Author:
Gunther Eysenbach, MD, MPH, FACMI
University Health Network
Centre for Global eHealth Innovation & Techna Institute
190 Elizabeth St
Toronto, ON, M4L3Y7
Canada
Phone: 1 416 7866970
Fax: 1 416 340 3595
Email: geysenba@uhnres.utoronto.ca

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2012/1/e7

Abstract

Background: Citations in peer-reviewed articles and the impact factor are generally accepted measures of scientific impact. Web 2.0 tools such as Twitter, blogs or social bookmarking tools provide the possibility to construct innovative article-level or journal-level metrics to gauge impact and influence. However, the relationship of the these new metrics to traditional metrics such as citations is not known.

Objective: (1) To explore the feasibility of measuring social impact of and public attention to scholarly articles by analyzing buzz in social media, (2) to explore the dynamics, content, and timing of tweets relative to the publication of a scholarly article, and (3) to explore whether these metrics are sensitive and specific enough to predict highly cited articles.

Methods: Between July 2008 and November 2011, all tweets containing links to articles in the Journal of Medical Internet Research (JMIR) were mined. For a subset of 1573 tweets about 55 articles published between issues 3/2009 and 2/2010, different metrics of social media impact were calculated and compared against subsequent citation data from Scopus and Google Scholar 17 to 29 months later. A heuristic to predict the top-cited articles in each issue through tweet metrics was validated.

Results: A total of 4208 tweets cited 286 distinct JMIR articles. The distribution of tweets over the first 30 days after article publication followed a power law (Zipf, Bradford, or Pareto distribution), with most tweets sent on the day when an article was published (1458/3318, 43.94% of all tweets in a 60-day period) or on the following day (528/3318, 15.9%), followed by a rapid decay. The Pearson correlations between tweetations and citations were moderate and statistically significant, with correlation coefficients ranging from .42 to .72 for the log-transformed Google Scholar citations, but were less clear for Scopus citations and rank correlations. A linear multivariate model with time and tweets as significant predictors ($P < .001$) could explain 27% of the variation of citations. Highly tweeted articles were 11 times more likely to be highly cited than less-tweeted articles (9/12 or 75% of highly tweeted article were highly cited, while only 3/43 or 7% of less-tweeted articles were highly cited; rate ratio 0.75/0.07 = 10.75, 95% confidence interval, 3.4–33.6). Top-cited articles can be predicted from top-tweeted articles with 93% specificity and 75% sensitivity.

Conclusions: Tweets can predict highly cited articles within the first 3 days of article publication. Social media activity either increases citations or reflects the underlying qualities of the article that also predict citations, but the true use of these metrics is to measure the distinct concept of social impact. Social impact measures based on tweets are proposed to complement traditional

XSL•FO
RenderX

citation metrics. The proposed twimpact factor may be a useful and timely metric to measure uptake of research findings and to filter research findings resonating with the public in real time.

(J Med Internet Res 2011;13(4):e123) doi:10.2196/jmir.2012

KEYWORDS

bibliometrics; blogging; periodicals as topic; peer-review; publishing; social media analytics; scientometrics; infodemiology; infometrics; reproducibility of results; medicine 2.0; power law; Twitter

Introduction

Scientists, research organizations, and funding agencies require metrics to measure the impact of research. Citations in peer-reviewed articles referencing other articles are a widely accepted measure of scientific impact. Citations are the basis for metrics like the h-index [1] and its derivatives, which are used to evaluate the productivity and impact of individual researchers, or the impact factor, which is used to evaluate the scientific impact of journals [2]. However, citations as a metric have various disadvantages, including the fact that they take a very long time to accumulate. They are also difficult to obtain (in an environment where the majority of research is still not open access) and are often available only in proprietary databases; thus, these metrics are not necessarily transparent or reproducible. For example, the h-index of a researcher varies widely depending on the database used to calculate it, and calculation of the journal impact factor has been criticized for not being transparent [3,4]. Finally, citations measure only uptake within and impact on the scientific community, not, for example, impact on or dissemination among knowledge users (policy makers, patients, and the general public). While this may be desirable for some use cases, other applications and stakeholders require a broader definition of impact. Concepts such as impact on society, social impact, real-world impact, knowledge translation, and uptake by the public should be part of every research assessment exercise but are notoriously difficult to measure [5]. Tools such as questionnaires applied to publications have been suggested to measure the "societal impact factor" [6], but it is unclear whether these instruments, which require manual data collection, are scalable to a large number of publications.

In this paper I propose new metrics and a new source of data—Twitter—that could be used to measure social impact, complementing traditional citation analyses, pilot tested and illustrated on a set of articles from the *Journal of Medical Internet Research* (JMIR).

Web citation analysis has previously been used to measure the extent to which articles or ideas are mentioned on the Web [7]. For example, Vaughan and colleagues have shown relationships between link metrics [8] or Web mentionings [9,10] and traditional impact metrics. Kousha and colleagues propose an "integrated online impact indicator" [11], which combines a range of online sources into one indicator for impact on the Web, including course reading lists, Google blogs, PowerPoint presentations [12], and Google Books [13].

Web 2.0 tools such as Twitter and blogs, as well as social bookmarking tools and Web-based reference management tools such as CiteULike and Mendeley, provide the opportunity to gather novel metrics from other sources that provide data in a structured format, accessible through application programming interfaces (APIs) [14,15]. These metrics—sometimes called altmetrics [16] or (in a broader context) infodemiology metrics [17,18]—can be used to gauge concepts such as popularity, buzz, social impact, or uptake of new information. The underlying common idea is that scientists and the public leave digital traces on the Internet when searching for or using information, and the Web has "made measurable what was previously immeasurable," [18] which is the demand for or use of specific information, and dissemination of information, as it propagates through networks. Infodemiology is an emerging area of science with applications in public health [17,18] and a wide range of other areas [19]—it has, for example, been shown that search engine queries predict influenza [20,21], that tweets during the H1N1 pandemic correlated with incidence rates [22], and that tweets about a movie accurately predict its box-office success before the movie is even released [23].

In analogy to the applications for public health 2.0 [17], economics, and other areas [19], there is an obvious application of infodemiology or infoveillance for *scientometrics 2.0* [24], which is to study the buzz around scientific publications to measure or even predict the impact of research.

The field of social media-based scientometrics (altmetrics, infodemiology metrics) is in its infancy, and many open questions need to be addressed. It may be that these new metrics measure completely different concepts that are not correlated with other traditional metrics such as citations, but it may also be that important publications in the scholarly literature first lead to a measurable buzz within the blogosphere (and other Web 2.0 venues) before, years later, the buzz is also reflected in increased citations and/or policy changes and social impact.

Specific questions include the following. (1) How can buzz be measured? (2) When (in relation to the publication of an article) and how long should we measure it? (3) If we can measure something, how are the metrics related to traditional metrics such as citations, and is the buzz sensitive enough to predict increased citations? (It should be noted that prediction of citations is not necessarily the end goal, and that lack of correlation is not necessarily a failure, because it is clear that these metrics add a new dimension of measuring impact.)

There is a dearth of empirical data exploring and showing such relationships, which would be seminal to develop the field of social media-based scientometrics. While it has been shown that scholars cite on Twitter and reasons for scholars to do so have been explored [24], little is known how—on an article- or

XSL•FO
RenderX

journal-level—publications attract tweets, and whether meaningful metrics can be derived.

There is a small but quickly growing body of literature focusing on Twitter for use in scholarship [24-29]. Most papers focus on analyzing Twitter streams collected during conferences [25-27], while little or no evidence is available on a journal level. The Public Library of Science (PLoS) journals make available some article-level impact metrics, which scholars have started to analyze [30], but PLoS has only recently begun to count tweets.

At JMIR we started the current empirical, prospective study in 2008, at a time when few journal publishers or scholars thought about the potential of Twitter for analyzing impact. The goals of the current study were (1) to explore the content and characteristics of tweets discussing or mentioning research articles and their timing relative to the publication date of an article, (2) to identify suitable metrics to describe propagation of new evidence through social media networks, and (3) to explore how the proposed metrics correlate with traditional metrics of uptake within the scientific community (traditional citations).

Methods

JMIR Twitter Dataset and Tweetation Counts

JMIR is a leading, highly cited open access journal with a Thomson Reuters (formerly ISI) 3-year impact factor of 4.7 and 5-year impact factor of 5.0 (Journal Citation Reports, 2010). In July 2008, it was the first journal to start systematically mining tweets that mention its published articles, showing them in real time on the JMIR "Top Articles" Page (see Figure 1). Data are collected using the Twitter Search API.

For the purpose of this paper, I call a citation in a tweet (mentioning a journal article URL) a "tweetation", to distinguish it from a citation in a journal article (which is the metric I compared tweetations against). As 1 tweet can have multiple tweetations (a tweet containing multiple different URLs citing different articles), the number of tweetations is not necessarily identical to the number of tweets, although in our sample a tweet with multiple tweetations was very rare, so that I sometimes use *tweets* and *tweetations* interchangeably. Only tweets with URLs linking directly to the journal article are captured—that is, links to newspaper articles mentioning published research in JMIR or links to JMIR articles that are not on the JMIR site (eg, instances in PubMed Central, or links to the digital object identifier [DOI] handle)—are not counted. Retweets of the same tweet or sending a modified tweet by other users would count as multiple tweetations, as would multiple tweets from the same user containing the same URL.

XSL•FO
RenderX

Figure 1. Top Articles ranking on the *Journal of Medical Internet Research* (JMIR) (sorted by most-tweeted articles in November 2011).



Top Articles (Article-Level Metrics)

OVERVIEW MOST VIEWED **MOST TWEETED** MOST PURCHASED MOST CITED

ALL TIME* **IN THE LAST MONTH** IN THE LAST SIX MONTHS IN THE LAST YEAR

RANK	ARTICLE		TWEETS	TWEETS PER MONTH[A]	TWEETS INFLUENCE FACTOR[B]
1	Public Access and Use of Health Research: An Exploratory Study of the National Institutes of Health (NIH) Public Access Policy Using Interviews and Surveys of Health Personnel *Jamie O'Keeffe, John Willinsky, Lauren Maggio* J Med Internet Res 2011;13(4):e97 (*Nov 21, 2011*)	HTML *free!* PDF *($22 pay-per-download, free for members [$4.92/mo.]*	51	395.35	120.00
2	Physician Response Time When Communicating With Patients Over the Internet *Per Egil Kummervold, Jan-Are K Johnsen* J Med Internet Res 2011;13(4):e79 (*Nov 1, 2011*)	HTML *free!* PDF *($22 pay-per-download, free for members [$4.92/mo.]*	33	42.62	38.00
3	Analysis of 4999 Online Physician Ratings Indicates That Most Patients Give Physicians a Favorable Rating *Bassam Kadry, Larry F Chu, Bayan Kadry, Danya Gammas, Alex Macario* J Med Internet Res 2011;13(4):e95 (*Nov 16, 2011*)	HTML *free!* PDF *($22 pay-per-download, free for members [$4.92/mo.]*	29	99.90	91.00
4	Hospital-Based Nurses' Perceptions of the Adoption of Web 2.0 Tools for Knowledge Sharing, Learning, Social Interaction and the Production of Collective Intelligence *Adela S.M. Lau* J Med Internet Res 2011;13(4):e92 (*Nov 11, 2011*)	HTML *free!* PDF *($22 pay-per-download, free for members [$4.92/mo.]*	27	59.79	75.00
5	Anxiety Online—A Virtual Clinic: Preliminary Outcomes Following Completion of Five Fully Automated Treatment Programs for Anxiety Disorders and Symptoms *Britt Klein, Denny Meyer, David William Austin, Michael Kyrios* J Med Internet Res 2011;13(4):e89 (*Nov 4, 2011*)	HTML *free!* PDF *($22 pay-per-download, free for members [$4.92/mo.]*	23	33.95	41.00
6	Using Virtual Reality to Provide Health Care Information to People With Intellectual Disabilities: Acceptability, Usability, and Potential Utility *Valerie Hall, Suzanne Conboy-Hill, Dave Taylor* J Med Internet Res 2011;13(4):e91 (*Nov 14, 2011*)	HTML *free!* PDF *($22 pay-per-download, free for members [$4.92/mo.]*	23	64.83	32.00
7	Effects of an eHealth Literacy Intervention for Older Adults	HTML *free!*	17	23.95	38.00

Citation Counts

Citation counts were harvested from Scopus and Google Scholar. The current study is based on citation counts obtained in November 2011, which is 17–29 months after the cited papers were published.

Analysis

For the tweets distribution analysis all tweets sent and archived by JMIR between July 24, 2008 and November 20, 2011 were included (Multimedia Appendix 1).

For the tweetation–citation correlation analysis, I included only tweets that referred to articles published in issue 3/2009 through issue 2/2010—that is, tweetations of all 55 articles published

XSL•FO
RenderX

between July 22, 2009 and June 30, 2010 (Multimedia Appendix 2). I chose this period because the tweetation rate for earlier articles was too sparse, and later articles did not have enough citations accumulated as of November 2011.

Pearson correlations on the raw and the log-transformed data, as well as the Spearman rank correlations, were calculated. Data were log transformed using the natural logarithm because tweetation and citation data are highly skewed. As the log of 0 is undefined, 1 was added to the counts of citations and tweetations.

For the categorical classification analysis (attempts to predict highly cited articles from highly tweeted articles), "highly cited" articles were defined as articles that were in the top 25th percentile of each issue (articles ranked by citation counts), and "highly tweeted" articles were defined as articles that were in the top 25th percentile of each issue (ranked by tweetations).

The analysis was stratified on a quarterly per-issue basis to adjust for time as a confounder, because the popularity of Twitter (and the number of JMIR followers) increased over time (older articles will have fewer tweets than newer articles), and because older articles will have more citations than more recent ones. Stratification by journal issue assures that the articles that were compared against each other were all published within the same quarter (3-month window).

In another analysis I included articles from all issues, but adjusted for time as a potential confounder by conducting a linear regression analysis, with the logarithm of citations as dependent variable, and time (days since publication of the earliest article in our dataset) and the logarithm of tweetations as independent variables.

Note that when article IDs are mentioned in this paper (see figures), these are part of the DOI; and each article can be identified by entering http://dx.doi.org/10.2196/jmir.{articleID} in a Web browser's address bar.

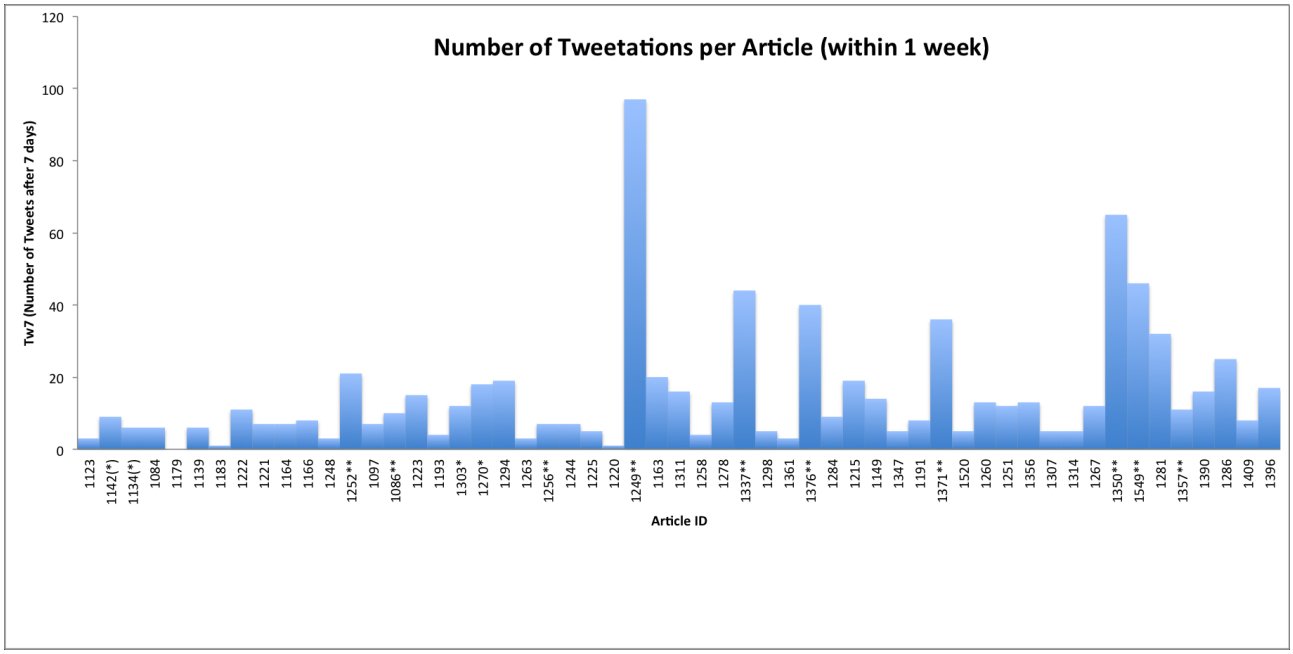
Results

Average Number of Tweets per Article

A total of 4208 tweetations were identified, which cited a total of 286 distinct JMIR articles, with each article receiving on average 14 tweetations (median 9). However, these averages should be interpreted with care, as JMIR has published articles since 1999 (560 articles in total). Among the 286 articles referenced in tweetations, there were many articles that were published before data collection began or before Twitter even existed. As these older articles receive only sporadic tweetations, the average and median are not reflective of more recent articles.

The 55 articles published in issues 3/2009–2/2010 received an average of 21.2 tweetations within 356 days after article publication (median 12, range 0–149), and 13.9 (median 8, range 0–96) tweetations within 7 days. Figure 2 shows the cumulative number of tweetations within 7 days (tw7) for these articles.

Figure 2. Number of tweetations within 7 days of article publication, per article ID. Asterisks next to article IDs denote that the article is top-cited (see also Figure 8): ** top 25th citation percentile within issue by both Scopus and Google Scholar citation counts * top 25th citation percentile according to Google Scholar only, (*) top 25th citation percentile according to Scopus only.



Tweet Dynamics

When, in relationship to the date of publication of an article, did the tweetations occur? Figure 3 shows the general distribution of all tweetations (n = 3318) that were sent within 60 days after publication of the article they are citing, by day. In this graph, day 0 refers to the day of article publication, day 1 is the following day, and so on; the left y-axis shows how many of the tweetations were sent on that day (tweet rate), as a proportion relative to all tweetations within a 60-day period; and the right y-axis (and red line) shows the cumulative proportion. The majority of tweets were sent on the day when an article was published (1458/3318, 43.9%) or on the following

XSL•FO
RenderX

day (528/3318, 15.9%). Only 5.9% (197/3318) of all tweetations are sent on the second day after publication, and the downward trend continues, until a little plateau between days 5 and 7 occurs (about 2% of all 60-day tweetations). There is a dip on days 8 and 9, which may be explained by the fact that, while JMIR publishes articles on different days of the week, Friday is slightly more prevalent, so days 8 and 9 would fall on the following weekend. After day 10 (66/3318, 2%) the rate of new tweetations declines rapidly.

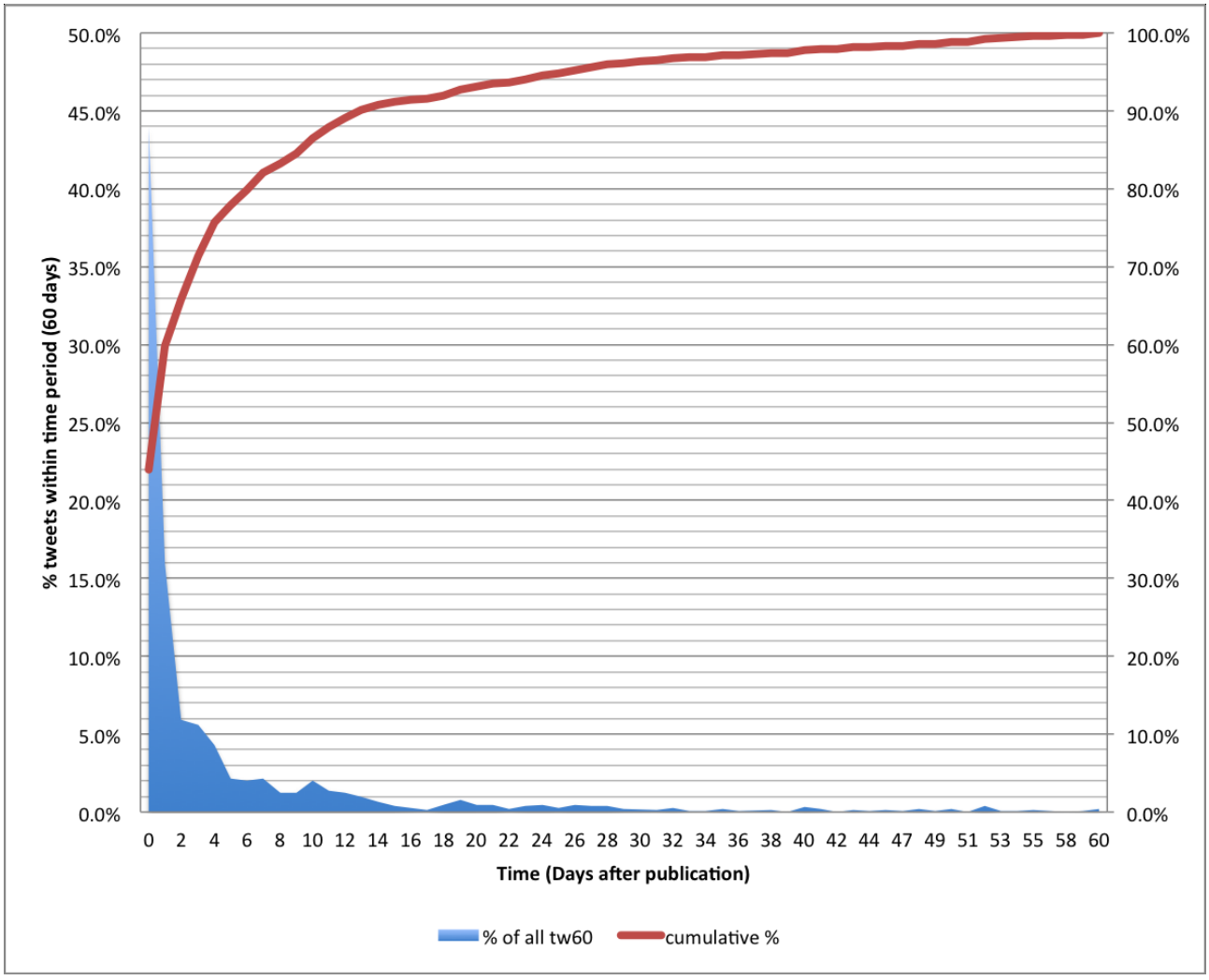
Figure 4 shows the same curve of new tweetations by day, but this time replotted with logarithmic horizontal and vertical axes. Now an interesting pattern emerges, showing a strong regularity: the tweetation distribution during the first 30 days on a log–log plot follows a straight line, which is indicative of a Pareto distribution, also known as Zipf's law or Bradford distribution, which are said to follow a power law [31]. In our sample, the number of tweetations per day after the article has been published during the first 30 days can be predicted by the formula $\ln(tw) = -1.53 * \ln(d) + 7.25$, where tw is number of new tweetations on day d, and d is days since publication (publication date = day 1).

This model has an excellent fit ($R^2 = .90$). While the intercept of this formula is not important (it is dependent on the total number of tweetations), the term –1.53 is called alpha or the exponent of the power law (slope of the linear curve in the log–log diagram).

We can divide the pattern in Figure 4 into two distinct phases: I call the first 30 days the "network propagation phase," where the new information is propagated through the Twitter social network. After 30 days, the network propagation phase gives way to what I call the "sporadic tweetation phase," where only sporadic mentionings of older articles and small clusters of localized outbreaks of information propagation occur.

Figure 5 shows the tweetation dynamics for all articles in JMIR issue 1/2010. Note that while Figure 4 shows the number of *new* tweetations per day (tweet rate, which is sharply declining), Figure 5 shows them in a cumulative manner. The figure illustrates how some articles attract tweets only on the first day, while some other articles continue to attract tweetations and are more widely retweeted. Incidentally, these are often articles that turn out to be highly cited, as shown in more detail below.

Figure 3. Tweetation dynamics. The blue, shaded area (left y-axis) shows the tweet rate (new tweetations per day, as a proportion of all tweetations during the first 60 days [tw60]). The red line (right y-axis) represents cumulative tweetations.



XSL•FO
RenderX

Figure 4. Tweetation dynamics over time on a log-log scale. All tweetations were categorized according to when, in relationship to the cited article publication date, they were tweeted (x-axis), with 1 being the day of article publication.

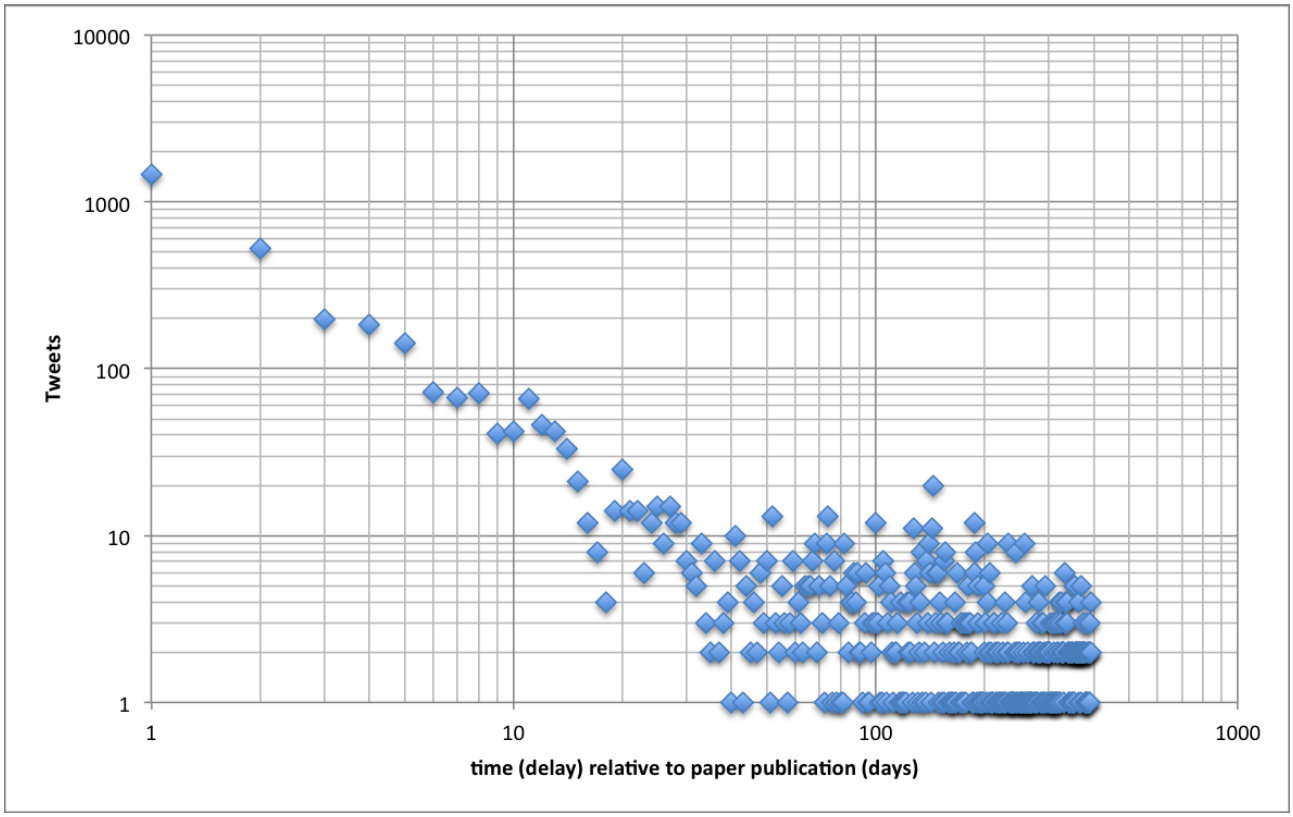
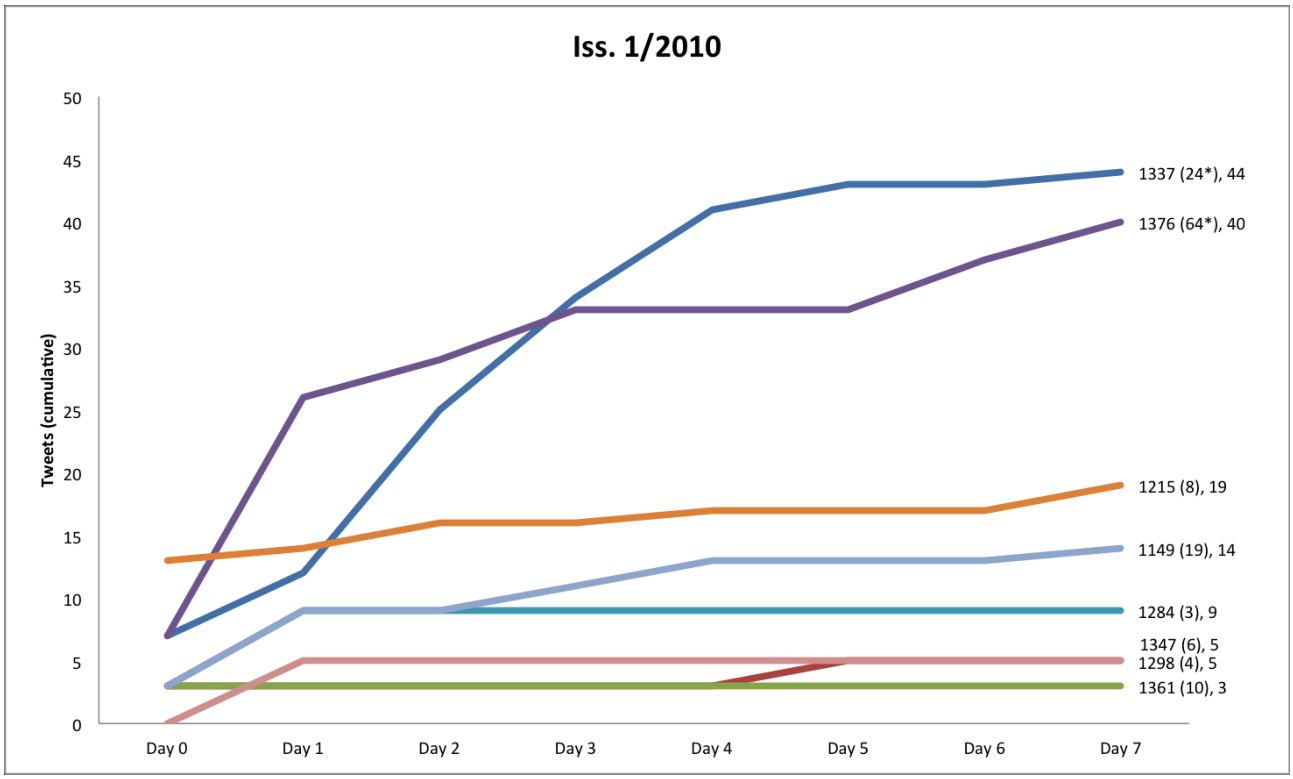


Figure 5. Tweetation dynamics in the first 7 days after article publication for one specific issue. The 4-digit number is the article identifier (last digits of the DOI), number in parentheses is the citation count (as per Google Scholar, November 2011), and the last number is the (cumulative) number of tweets on day 7 (tw7).



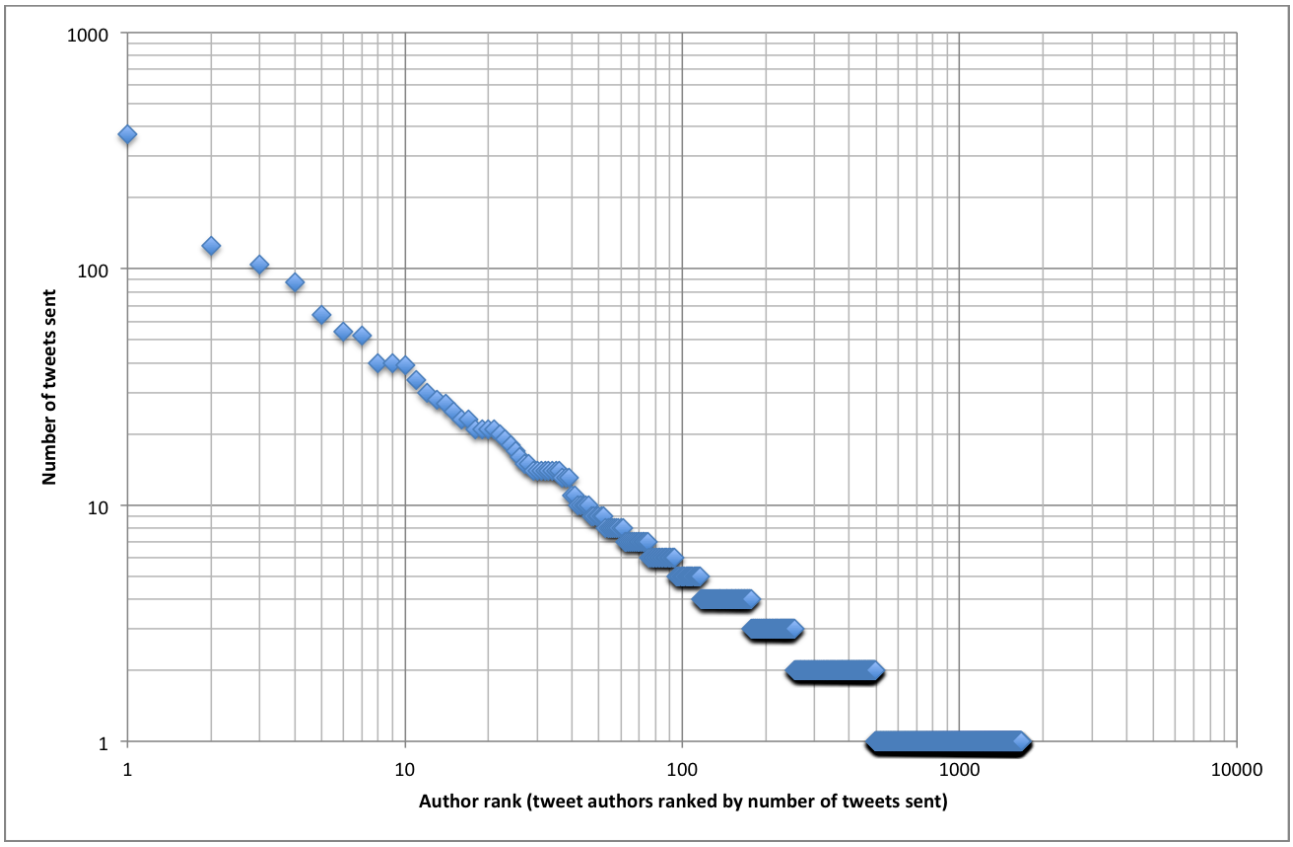
XSL•FO
RenderX

Other Regularities

There were other strong regularities of tweetations following power laws. Tweetations were sent from 1668 distinct Twitter accounts (tweet authors). The most tweetations (n = 370) were sent by @JMedInternetRes, JMIR's Twitter account. If we rank the accounts by the number of tweetations they sent and plot them against the number of tweetations for each account, the power law distribution shown in Figure 6 emerges. Half of all tweets (2105/4208, 50%) were sent by only 132 distinct tweet authors—that is, 8% of all tweet authors. The top 20% of the tweet authors (those ranked 1–334 by number of tweetations) accounted for 63.4% (2676/4208) of all tweetations. This uneven distribution of work is typical for Pareto distributions, an observation that is sometimes colloquially referred to as the 80/20 rule, where roughly 80% of the effects come from 20% of the causes.

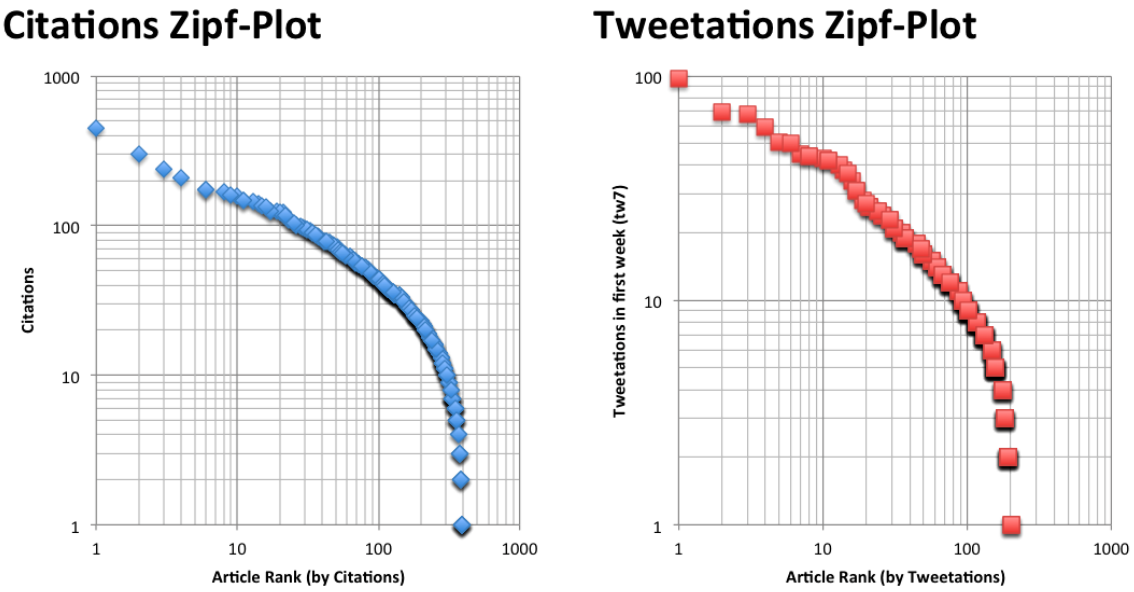
The third power law I looked at was where I expected it most, because this distribution is typically observed for citations and can be demonstrated in a Zipf plot, in which the number of citations of the *n*th most-cited paper is plotted versus the rank n (Figure 7, left). Tweetations follow a strikingly similar distribution (Figure 7, right).

Figure 6. Tweetation density by account. Each Twitter account is ranked by the number of tweetations sent and plotted by rank on the x-axis. The y-axis shows how many tweetations were sent by each ranked account. For example, the top Twitter account ranked number 1 (@JMedInternetRes) sent 370 tweetations. Note the linear pattern on a log-log scale, implying a power law.



XSL•FO
RenderX

Figure 7. Left: Zipf plot for JMIR articles 3/2000-12/2009 (n=405), with number of citations (y-axis) plotted against the ranked articles. Right: Zipf plot showing the number of tweetations in the first week (tw7) to all JMIR articles (n=206) published between April 3, 2009 and November 15, 2011 (y-axis) plotted against the ranked articles. For example, the top tweeted article got 97 tweetations, the 10th article got 43 tweetations, and the 102th ranked article got 9 tweetations.



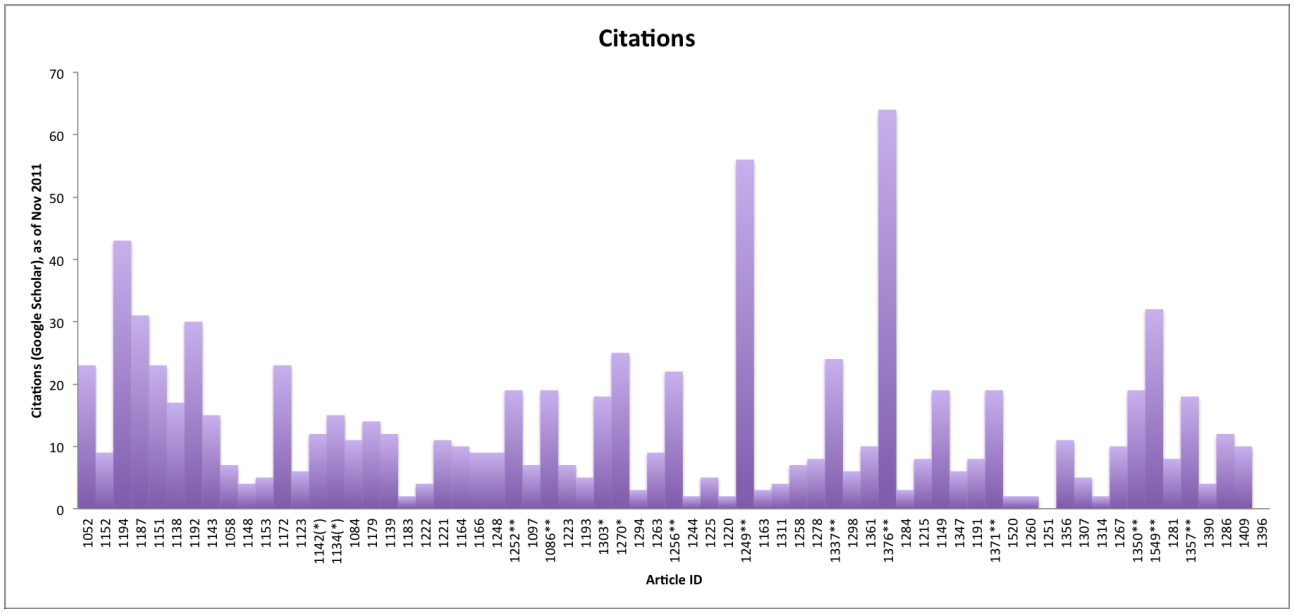
Citations

The 55 articles in our tweetations-versus-citations subset had an average of 7 citations on Scopus (median 4) and 13 citations on Google Scholar (median 9). Figure 8 shows the Google Scholar citation counts for all 55 articles included in the tweetation/citation analysis, as of November 2011.

First, the number of citations from Scopus were correlated with the number of citations from Google Scholar to test agreement between the two database sources. There was good agreement, with a Pearson correlation coefficient of .87 ($P < .001$) for the 55 articles. As Google Scholars' citation counts were higher and appeared more robust, most results presented here refer to Google Scholar citation counts, unless noted otherwise.

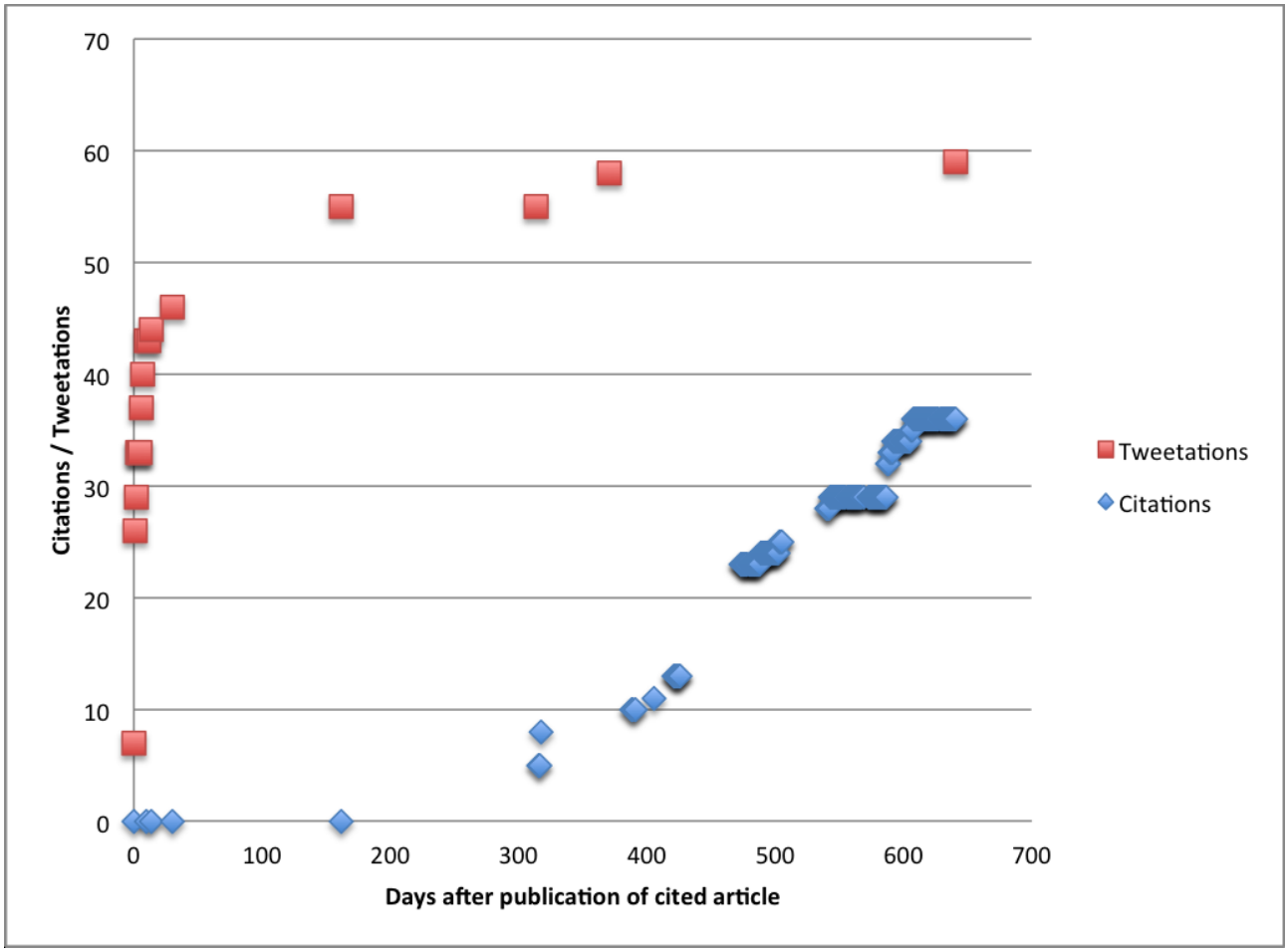
Figure 9 compares a typical citation and a tweetation curve, illustrating the very different dynamics in tweetations compared with citations in scholarly articles. While citations in scholarly articles begin to accumulate only about 1 year after the article is published, tweetations accumulate mainly within the first few days after publication.

Figure 8. Google Scholar citation counts for all articles published between issue 3/2009 and issue 2/2010. Top-cited articles (75th percentile) within each issue are marked ** (top cited according to Google Scholar and Scopus), * (Google Scholar only), or (*) (Scopus only).



XSL•FO
RenderX

Figure 9. Citation and tweetation dynamics of a highly cited (and highly tweeted) article [article ID 1376]; citations according to Scopus.



Correlation Between Tweetations and Citations

For each journal issue, I separately plotted scatterplots and calculated Pearson correlation coefficients of the raw count, the logs, and Spearman rank correlation coefficients, to establish the degree of correlation between citations and tweetations.

My primary tweetation metric was tw7 (cumulative number of tweetations 7 days after publication of the article, with day 0 being the publication date), a metric I also call *twimpact factor* or TWIF7 (see below).

The Pearson product moment correlation coefficients (*r*) for the raw citation versus tw7 tweetation counts were statistically significant on a 5% level for all journal issues, and ranged from .57 to .89 (Table 1). Pearson correlations between the logs of citations and logs of tweets, as well as Spearman rank correlation coefficients, were all statistically significant when articles across issues were combined, except for the rank correlation between Scopus citation counts and tweetations. When stratified by journal issue, the correlations for some issues were statistically significant for some computations, while for others they were not, perhaps due to a small sample size. Generally, the Google Scholar citations showed better correlations with tweetations than did Scopus citations (Table 1). The Spearman rank correlations (rank by citations versus rank by tw7) were statistically significant for only one issue, with rho = .51, P = .04 for issue 2/2010.

I also conducted analyses with other tweetation metrics (tw0, tw1, tw2, tw3, tw4, tw5, tw6, tw7, tw10, tw12, tw14, tw30, and tw365) and derived various metrics (tw365–tw7, ie, late-stage tweets; tw7–tw0, tw0/tw7 etc), which produced very similar correlation coefficients (data not shown).

XSL•FO
RenderX

Table 1. Correlation coefficients

	Issue (number of papers)									
	3/2009 (n = 19)		4/2009 (n = 11)		1/2010 (n = 8)		2/2010 (n = 17)		All (n = 55)	
	r or rho	*P* value	*r* or rho	*P* value	*r* or rho	*P* value	*r* or rho	*P* value	*r* or rho	*P* value
Pearson correlation (*r*)										
CitGo-Tweets[a]	.57**	.01	.89***	<.001	.76*	.03	.68**	.003	.69***	<.001
CitSc-Tweets[b]	.33	.17	.74**	.01	.65	.08	.51*	.04	.54***	<.001
logCitGo-logTweets	.42	.08	.51	.11	.72*	.045	.49*	.048	.39**	.004
logCitSc-logTweets	.03	.90	.41	.22	.53	.17	.47	.06	.31*	.02
Spearman rank correlation (rho)										
CitGo-Tweets	.42	.07	.14	.68	.61	.11	.51*	.04	.36**	.006
CitSc-Tweets	.06	.81	.11	.76	.44	.27	.42	.10	.22	.11

[a] Citation count according to Google Scholar (CitGo) versus tweetation count (tw7).

[b] Citation count according to Scopus (CitSc) versus tweetation count (tw7).

*P < .05, **P < .01, ***P < .001.

Multivariate Analysis

In a linear regression model I tried to predict the log of the number of Google Scholar citations from the log of the number of tweets and time (days since publication of the first article in the sample of 55 articles). The regression equation was log(cit + 1) = 0.467 * log(tw7 + 1) + –.001 * days + 0.817, where cit is the number of citations, and tw7 is the cumulative number of tweetations at day 7. Both independent variables were significant predictors (P < .001), and the model explained 27% of the variation of citations (R^2 = .27).

Binary Analysis

Based on the observation that tweets were sent primarily during the early days after publication, I hypothesized that tw7, the cumulative number of tweetations by day 7 (perhaps as early as day 3), could be used as a diagnostic test to predict highly cited articles. *Highly tweeted* and *highly cited* are defined as articles in the 75th–100th percentile of each journal issue; thus, the cut-off points on what constitutes highly tweeted or highly cited varied by issue (tweets: 11, 19, 34.8, 28.5; Google Scholar citations: 15, 9, 22.75, 15, for issues 3/2009, 4/2009, 1/2010, and 2/2010, respectively).

Table 2 is a 2 × 2 table categorizing articles into the four groups. Articles that were less frequently tweeted and not in the top-cited quartile are interpreted as true negatives (tn, lower left quadrant in Figure 10 and Table 2). Articles that were highly tweeted and highly cited are true positives (tp, upper right quadrant in Figure 10 and Table 2). Articles that were highly tweeted but not highly cited fall into the upper left quadrant and are referred to as false positives (fp). Finally, articles that were not highly tweeted but highly cited are false negatives (fn).

Using tweetation status (highly versus less tweeted) as a predictive test for citation status, this test identified 40 out of the 43 not highly cited articles, which translates to a 93% specificity (true-negative rate, tn/[tn + fp], 40/43). The test was able to correctly identify 9 out of the 12 highly cited papers, which corresponds to a 75% sensitivity (tp/[tp + fn], 9/12). Another way to express these results is to say that the positive predictive value (tp/[tp + fp]) or precision is 75%, meaning that if an article is highly tweeted (tests positive for social media impact), then there is a 75% likelihood that the article ends up in the top quartile of all articles of an issue, ranked by citations. The negative predictive value (tn/[tn + fn]) is 93% (40/43), meaning that if an article was not highly tweeted (tests negative for social media impact), then there is only a 7% (3/43) chance that it will fall into the top 25% of cited articles. Yet another way to express these results is to say that highly tweeted articles are almost 11 times more likely than less tweeted articles to be highly cited (9/12, 75% highly tweeted article are highly cited, while only 3/43, 7% of the less tweeted articles are highly cited; rate ratio 0.75/0.07 = 10.75, 95% confidence interval, 3.4–33.6).

There was a highly statistically significant association between citation status and tweetation status (Fisher exact test, P < .001).

I repeated this analysis for a range of different metrics such as tw*n* (cumulative number of tweetations after *n* days, with n = 0, 1–10, 12, 14, 30, or 365), and the number of late-response tweetations tw365–tw7. Starting on day 3 (tw3), the heuristic started to identify the same top-tweeted articles as tw7, indicating that the test is predictive as early as 3 days after publication. Choosing later days (letting tweetations accumulate for more than 7 days) or the late-response tweetations did not improve the test results (data not shown).

XSL•FO
RenderX

Figure 10. Correlations between citations in November 2011 (Google Scholar) and the cumulative number of early tweets by day 7 (tw7). Note the logarithmic scale. Articles with 0 tweets or 0 citations are not displayed here, because the log of 0 is not defined. However, conceptually they all fall into the lower left quadrant.

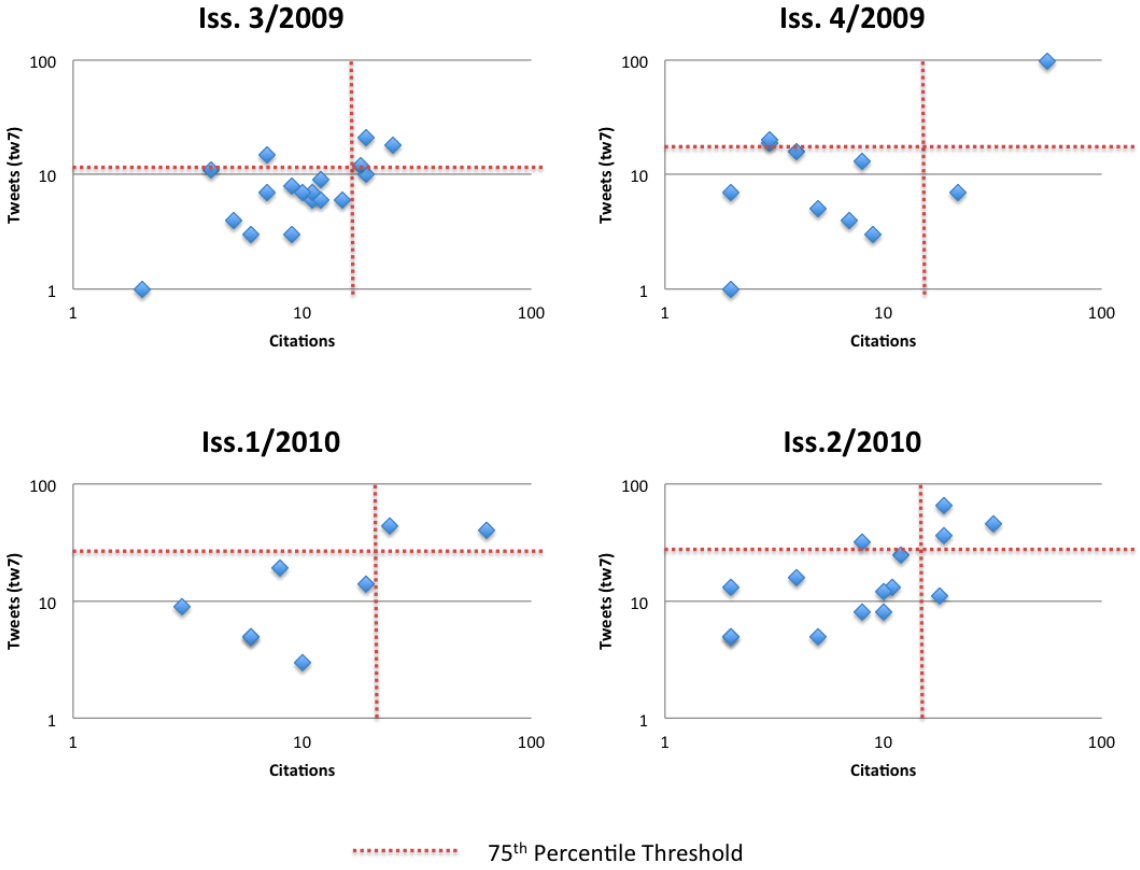


Table 2. 2 × 2 table using top-tweeted articles as a predictor for top-cited articles

	Less cited (bottom 75%) n = 43	Highly cited (top 25%) n = 12
Highly tweeted (top 25%) n = 12	fp[a] (n = 3) [Article ID 1223, 1163, 1281]	tp[b] (n = 9) [Article ID 1252, 1303, 1270, 1249, 1337, 1376, 1371, 1350, 1549]
Less tweeted (bottom 75%) n = 43	tn[c] (n = 40)	fn[d] (n = 3) [Article ID 1086, 1256, 1357]

[a] False positives.

[b] True positives.

[c] True negatives.

[d] False negatives.

Proposed Twitter-Based Metrics for Social Impact

The research reported here focuses on articles from one journal. However, I suggest that the metrics introduced here should be useful to measure the impact any article (or collections or sets of articles) has on Twitter, to gauge how much attention users pay to the topic of an article, to measure how the question and/or conclusions resonate with Twitter users, and ultimately to use them as proxies for social impact. Although I use Twitter as an example here, these metrics can be used in other social media (eg, Facebook status updates). The metrics presented here can also be generalized and applied to measure the impact of any issue (not just scholarly articles but, for example, current events and newspaper articles) on a social media user population.

Twimpact Factor (eg, tw7)

Using raw tweetation counts to compare the impact of different articles with each other is problematic, because the number of tweetations is a function of time since publication. Although the data suggest that after an initial period of 30 days tweetations usually occur only sporadically, the raw number of tweets should not be used when comparing articles with each other if they have been published on different dates. An average tweetation count per month since publication is possible to calculate (and

XSL•FO
RenderX

is currently displayed on the JMIR Top Articles webpage, see Figure 1), but due to the highly skewed power law distribution, this average will always favor articles that have been published recently (within the last month).

I therefore propose to use (and have used in this paper) the twimpact factor tw*n* as a metric for immediate impact in social media, which is defined as the cumulative number of tweetations within *n* days after publication (eg, tw7 means total number of tweetations after n = 7 days). *Tweetations* can be replaced by *URL mentionings* if we apply this metric to other social media (URL being the URL or set of URLs of a specific article).

As a standard twimpact factor metric for an article on Twitter, I suggest (and JMIR will use in the future) tw7—that is, the absolute, cumulative number of tweetations an article receives by day *7* after publication (the day of publication is referred to as day 0). This is also a very practical metric: using a relatively short period of time makes the twimpact factor easier to compute, as the Twitter stream needs to be monitored for only 7 days.

I have shown that the number of new tweetations drops off rapidly after publication, even for the most highly cited papers. The immediate social media response is highly correlated with the later social media response; therefore, it is likely that the late response can be ignored. An even shorter period of time (3 days), tw3, was already sufficient in the sample to discriminate between highly cited and less cited articles, but I suggest a standard *n* of 7, which has the advantage that it always includes a weekend; thus, journal articles published toward the end of the week are less penalized for the weekend effect.

Any article, but also a collection of articles, can have a twimpact factor (eg, on a journal or issue level). JMIR is now monitoring the *collective twimpact factor* ctw*n*/*m* for each journal issue (where *n* is the number of days after publication tweetations accumulate, and *m* is the percentile), eg, ctw7/50 is the median (50^{th} percentile) of tw7 for all articles in the set. The ctw7/75 for JMIR issue 2/2010 is 29, meaning that the top 25% most-tweeted articles in issue 2/2010 were tweeted more than 29 times during the first week. We prefer to report the 75^{th} percentile instead of the mean or median (ctw7/50) because of the power distribution and because it seems a useful cut-off point to predict top-cited articles. At least in our sample, the practical meaning of the collective twimpact factor ctw7/75 is that articles with a tw7 greater than the ctw7/75 of a journal issue have a 75% likelihood of being top-cited (ending up in the top quartile of all articles of an issue, ranked by citations).

Note that the twimpact factor is an absolute measure counting tweetations; thus, just like for the journal impact factor, caveats apply. First, it is highly subject specific, so if comparisons are made between journals or even articles from the same journal, they should be made within a narrow subject category. An article on social media will more likely than an article about molecular biology be picked up by social media. Although within a specific field the twimpact factor may predict citations (predict which article is more likely to be highly cited), it would not be legitimate to compare the twimpact factor of an article on social media with a twimpact factor of an article about molecular biology, and conclude that the social media article will be more likely cited.

Second, similar to the caveat that journal impact factors should not be compared across different years, as the total number of citations is constantly growing, only articles that are published in a similar timeframe should be compared with each other (perhaps even 1 year is too long; thus, we made comparisons on a quarterly within-issue level). This is because both the number of Twitter users and the number of followers of a journal grow over time.

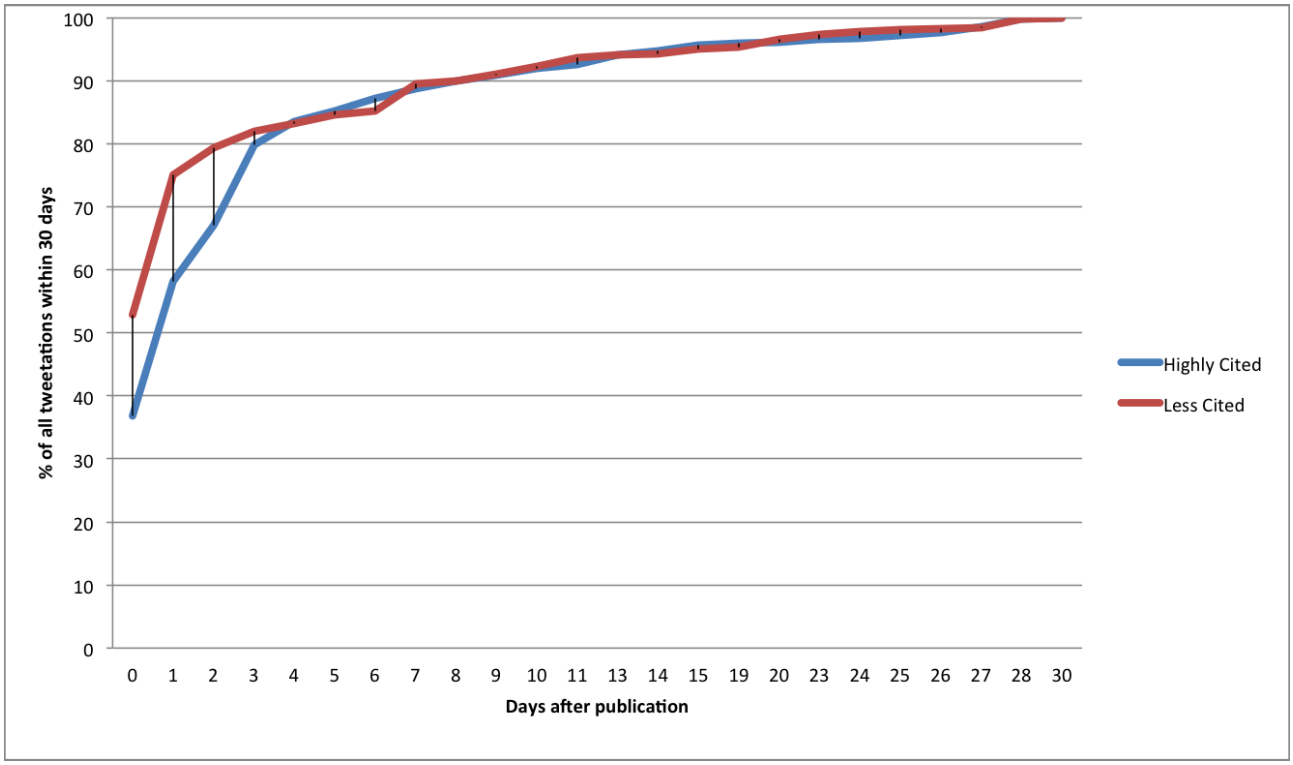
Tweeted Half-Life

The tweeted half-life (THL*n*) is defined as the point in time after publication by which half of all tweetations of that article within the first *n* days occur. As *n* I have used 30 days—that is, as the denominator I chose the total cumulative number of tweets within a 30-day period following the publication date. The THL*n* is the day when cumulatively half of these tweetations have occurred.

In our sample, the THL*n* for the less-cited articles was 0 (53% of the tweets were tweeted on day 0), while the THL*n* of highly cited articles was 1 (on day 0, 37% of all tweetations occurred, while on day 1, 21% occurred, in total 58% by day 1). Figure 11 illustrates this. It may at first seem surprising that less-cited articles appear to show a quicker and proportionally higher response on the first days, but it should be kept in mind that the *absolute* counts of tweetations for more highly cited articles are higher than for the less-cited articles. Low-impact articles are tweeted and retweeted mainly on day 0 and day 1. Highly cited articles continue to be retweeted widely, which depresses the *relative* proportion of tweetations on days 0–3.

XSL•FO
RenderX

Figure 11. Tweetation curves: cumulative tweetations (tw*n*), as a proportion of all tweetations sent within 30 days.



Twindex

As a final metric I propose (and JMIR will use) the twindex (tweetation index), which is a metric ranging from 0 to 100 indicating the relative standing of an article compared to other articles. I define the twindex$_7$ of specific article as the rank percentile of this article when all articles (the specific article and articles from a comparator group) are ranked by the twimpact factor tw7. The comparator articles should be similar articles published in a similar time window (eg, other articles in the same issue, or the 19 articles published previously in the same journal). If an article has the highest twimpact factor tw7 among its comparator articles, it has a twindex of 100. If it has the lowest twimpact factor, it has a twindex of 0. In this study, articles with a twindex > 75 often also turned out to be the most-cited one.

Discussion

Principal Findings

To my knowledge, this is the first systematic, prospective, longitudinal article- and journal-level investigation of how mentionings (citations or tweetations) of scholarly articles in social media accumulate over time. It is also the first study correlating *altmetrics* to subsequent citations. I have discovered important regularities that will be very useful for others interested in applying and developing social media-based impact metrics, not only in the context of scientometrics.

This paper shows that buzz in the blogosphere is measurable, and that metrics can be derived that are somewhat correlated with citations. Citations from Google Scholar seem more closely correlated with tweetations than are citations from Scopus, which likely reflects the fact that Google Scholar includes a wider range of citing sources, especially from nonjournal documents [32]. The Spearman rank correlations are poorer than Pearson correlations, probably because among the less-tweeted articles tweetations are sparse, and often as few as 1 or 2 tweetations make a difference on the ranking of an article. The correlation is, however, strong enough that we can make surprisingly accurate binary predictions along the lines that highly tweeted articles are 11 times more likely to end up being highly cited.

Correlation is not causation, and it harder to decide whether extra citations are a *result* of the social media buzz, or whether it is the underlying quality of an article or newsworthiness that drives both the buzz and the citations—it is likely a combination of both. It is not inconceivable that exposure on Twitter leads to a few extra citations: social media are often used by scientists "to catch useful citations...scholars might not otherwise be exposed to" [24], and many scientists see the value of Twitter in being a constant live literature alert service crowdsourced from peers. Tweets contain hyperlinks to articles, and hyperlinks may affect the ranking in search engines such as Google and increase the visibility for researchers.

Limitations of Twitter-Based Metrics

I suggest tweetations, twindex, and twimpact factor as metrics, which JMIR will publish and promote. These should be primarily seen as metrics for social impact (buzz, attentiveness, or popularity) and as a tool for researchers, journal editors, journalists, and the general public to filter and identify hot topics. Attentiveness to issues is a prerequisite for social change [33,34], and tweets are a useful metric to measure attentiveness to a specific scholarly publication. The data presented here also show that social impact is somewhat correlated with scientific impact, but there are important caveats. The correlation is far

XSL•FO
RenderX

from perfect (as one can expect), and the *complementary* nature of the metrics needs to be stressed (as an aside, the complementary nature is also why the term altmetrics is not favored by this author—these metrics are probably not an *alternative*, but a *complement* to traditional citations). Popularity—which is one dimension of what tweet metrics are measuring—is an extremely useful (and revenue-predicting) measure for commercial enterprises such as the entertainment industry, but there are enormous pitfalls to applying metrics of popularity to health and science, if they are not qualified by or complemented with other metrics. While for funding organizations, journal editors, and research organizations it may be very valuable to know which topics resonate with the public (are popular and paid attention to), even though they did not receive a lot of citations (the articles in the false-positive group), there is a real danger that research topics or findings that are not trendy enough to resonate with the Twitter population—for example, research affecting disadvantaged populations that are not represented on Twitter—are marginalized. It is interesting that one of the false negatives (many citations, but few tweetations) included a paper dealing with a low-income elderly population [Article ID 1256]—exactly the population that is underrepresented on Twitter. On the other hand, publications that are "sexy," trendy, or funny may receive huge exposure on Twitter, but may (or may not) have limited scientific value (a concept that is also not always accurately measured by citations).

Still, as mentioned earlier, there is enormous potential value for funding organizations, editors, and academic institutions to monitor these data, and to pay attention particularly to the false positives (high tweetations, low citations), as they may point to topics or questions that should perhaps be paid attention to. In our sample, the 3 articles that were highly tweeted but not highly cited (false positives; Article IDs 1223, 1163, 1281) all had a patient side to them, and consumers may have been the source of tweetations. Infoveillance of social media can be seen as a tool for public engagement in the discourse on what constitutes "important" research.

Finally, it must be acknowledged that there are journal-specific confounders at work that may limit the use of twimpact metrics, in particular if different journals are compared with each other (which is not currently done, but may be a future scenario). Journals cater to different communities and social networks, and when comparing how information propagates through online social networks, we may be measuring the structure of these networks and the attributes of these communities, rather than the attributes of the information itself. In other words, the number of tweetations is not a function of the intrinsic properties of the research article alone; rather, it is also influenced by factors related to the journal or venue it appears in, the community built around the journal, and how the scholarly information is marketed by the journal. But then again, the same is true for citations.

Limitations of This Study

While the results and metrics presented here are probably pivotal to paving the way to a new field of social media-based impact metrics, and while JMIR will increasingly use these approaches, the biggest question is whether our results and methods can be applied to other journals. JMIR is an ideal journal on which to experiment with altmetrics because it has a relatively high impact factor (ie, many traditional citation events) and—as a journal *about* the Internet and social media—it has a sophisticated readership that is generally ahead of the curve in adopting Web 2.0 tools. However, this also limits the generalizability of these results: what works for this journal may not work for other journals, in particular journals that are rarely cited (low impact factor) and that do not have an active Twitter user base. JMIR is a journal about information technology, and its readers may be more familiar with social media than readers of other journals are. Journals that publish non-Internet-related articles have probably far lower tweetation rates per article, and it is also less likely that people tweet about articles that are not open access. In fact, it has been argued that one key advantage of open access is that it facilitates knowledge dissemination among nonresearch users [35], and it is unlikely that articles from lower-impact subscription-based journals that are not accessible to a large number of users attract similar levels of tweetations. On the other hand, if tweetations about papers in subscription-based journals appear (eg, high-impact journals such as *Science* or *Nature*), it is likely that they were tweeted by expert users (scientist) who have access to the article; hence, they may be even more predictive for citations, because the general public is not (or to a lesser degree) part of the conversation.

The results presented here should be confirmed with tweets about other journals, as well as with future JMIR articles, and our group is currently conducting comparative analyses with other datasets. The hypothesis is that the results can be replicated for other journals as long as there is a large enough Twitter user base.

There are further, JMIR-specific caveats. First, as shown in Figure 1, JMIR ranks the top-tweeted articles on its website, and also sends out automatic tweets whenever a new article enters the top 10 in any of the monthly categories; both may have reinforced and amplified the response from Twitter users. Also, tweetations are a metric of the social media response; hence, the social media strategy of a journal likely has an impact on the results. Journals with an active social media presence and tweet alerts such as JMIR will have a higher uptake. JMIR followers have to click on only one button to retweet or modify these alerts (*seed tweets*). Journals that do not send out alerts for each article may have very different tweetation characteristics (eg, more late-stage tweetations). Further, the tweetation characteristics and rates are almost certainly influenced by the number of followers a journal has (JMIR currently has over 1000 followers) and, even more so, by lists and Twitter bots redistributing content to specific communities.

Researchers interested in using this new method and metric to compare different journals with each other should also be aware that the timing and frequency of article publication probably influence tweetation dynamics and rates (and may affect the strength of correlation between tweetations and citations). JMIR publishes articles as soon as they are ready, on different workdays of the week. As people tweet less during the weekend, the tweetation curve shown in Figure 2 may look slightly different for journals that always publish on Mondays (the

XSL•FO
RenderX

drop-off may be less pronounced), compared with a journal that publishes always on a Friday (here, the drop-off may be more pronounced). However, the tw7 metric (cumulative tweetations over the course of a week) is probably robust enough to compare journals with different publication schedules. Seasonal effects are also evident. For example, issue 5/2010 (not shown and not included in our analysis) was a theme issue published shortly before Christmas, and in this issue all articles were published at once rather than spread out over multiple days; as a result, articles in this theme issue had very low tweetation rates.

The current report does not include a systematic qualitative analysis of tweet contents. However, a cursory scan through all the tweets suggests that the vast majority of tweets simply contained variants of the article title or the key conclusion, and rarely contained explicit positive sentiments (such as "Great article!") or—even less common—negative sentiments (such as "questionable methods"—I have not seen any examples of the latter). This may be because the mere act of (re)tweeting an article is often an implicit endorsement or recommendation with which readers express their interest in and enthusiasm about a specific topic, support the research question and/or conclusion, or simply want to bring the article to the attention of their followers. Additional comments are not necessarily required to express this implicit endorsement. Also, with most tweets occurring on the day of publication, few readers will actually have had time to carefully read and appraise the entire paper beyond the title and perhaps abstract. While we originally thought of doing an automated sentiment analysis, the sparse nature of comments did not make this approach seem promising to elicit more specific data, although future studies using journals or articles with a high number of tweetations may want to take a close look at this question.

Future studies may also want to try to increase the specificity and sensitivity by focussing on specific types of twitter users, or taking into account the network structure and relative influence of the tweetation authors. JMIR publishes a tweets influence factor on its "Top Articles" Page (see Figure 1), which takes into account not only the number of tweets, but also the influence of the users who sent these tweets. The influence of users can be computed by the number of their followers and/or how often their tweets are retweeted, and more research is required to establish if these secondary metrics elicit additional information or are already reflected in the raw tweetation counts.

Another limitation is that the present analysis took into account first-order tweetations only. Tweets may contain links to blogs that in turn talk about articles, or may contain links to news articles that report on new research findings (second-order tweetations). According to Priem and Costello, about 50% are second-order tweetations [24]. This analysis did not capture these, as our tool strictly looks at tweetations with direct links to JMIR articles. We also did not capture links to other sites where JMIR articles may be hosted, including PubMed, PubMed Central, or DOI resolvers. Finally, twitter users commonly use URL shorteners, and while we retrieved some shortened URLs (by URL shorteners such as bit.ly), we may not have captured tweetations where the URL was shortened by less common shorteners. Thus, the true total number of tweetations was likely higher than what is reported here. On the other hand, there is no reason to believe that not counting these tweetations would introduce a bias.

In the current analysis each unique tweet was counted as 1 tweetation. Thus, multiple tweets sent by the same user about the same article would have been counted multiple times. This is not a problem in the current analysis, because multiple tweets with the same URL from the same user were quite rare. However, it is theoretically possible that—especially if tweetations become a more common method to rank and filter articles—authors may start to "game" the system by sending multiple tweets about their own article to create more exposure for their articles. Thus, for any use case with serious implications for authors (eg, if tweetations become a more accepted and common early metrics for social impact), a tweetation should be defined as an URL mentioned by a distinct unique user.

Conclusions

It is a fascinating and compelling finding that the collective intelligence of Twitter users can, within limitations, predict citations, which normally take years to accumulate.

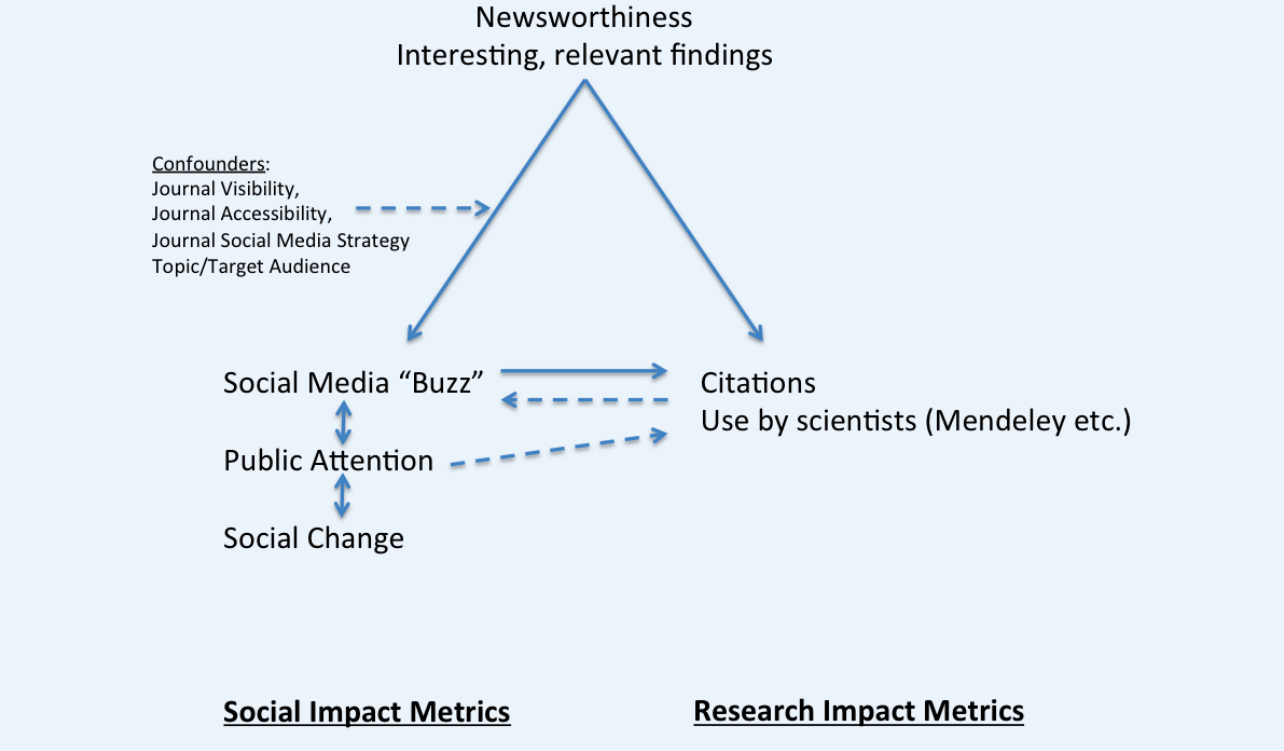
It should be stressed again that one should neither expect nor hope for perfect correlation. Tweetations should be primarily seen as a metric for social impact and knowledge translation (how quickly new knowledge is taken up by the public) as well as a metric to measure public interest in a specific topic (what the public is paying attention to), while citations are primarily a metric for scholarly impact. Both are somewhat *correlated*, as shown here, but tweetations and citations measure different concepts, and measure uptake by or interest of different audiences (Figure 12). The correlation and mutual interaction between these audiences is illustrated in Figure 12 with bidirectional arrows, which point from "social media buzz" to "citations" (scientists being influenced by social media buzz), and from "use by scientists" to "social media buzz" (scientists creating buzz on Twitter), illustrating the mutual influence of these audiences and metrics.

So if not (primarily) as a proxy or early indicators for citations, how should or could tweetations be used? What are the use cases?

First, social media metrics can be easily used by scholars, institutions, and journals to monitor the overall impact of research in a timely manner, keeping in mind the caveats and limitations listed above. Second, these metrics could be used to evaluate different methods of knowledge dissemination. One could design studies where different methods of promoting an article (or other URLs, for example public health intervention websites) are evaluated, with the twimpact factor as an outcome measure. Third, social media impact metrics can also be used as a filter to direct users to research articles that the public or research communities are paying attention to. A website displaying real-time social impact metrics such as twimpact factors of current research articles may be useful for a wide range of potential audiences, including journalists, journal editors, researchers, public health officials, and patients, to direct them to topics and research that resonate with the public.

XSL•FO
RenderX

Figure 12. Model of the relationship between social impact and research impact metrics.



A Standing Call for Papers

More research is required to assess the robustness of these social media metrics and their ability to detect signals among the noise of social media chatter, for scientometric purposes or other use cases in infodemiologic research. As mentioned earlier, the metrics and regularities presented here not only have applications for scientometrics, but also may be used to measure the dynamics and "half-life" of other issues or events discussed on Twitter or social media in general.

To stimulate and encourage innovation, research and development in this area, JMIR hereby issues a standing call for papers, welcoming empirical and viewpoint papers on the broad topic of infometrics or infodemiology metrics (or altmetrics, in the context of scientometrics), in particular with concrete use cases and data from health-related fields or journals. We look forward to publishing more research on what we feel are important methodological foundations for exploiting crowdsourcing and collective intelligence themes within the field of Internet research and science 2.0.

Acknowledgments

The author wishes to thank Phil Cairns for doing the coding work, Cynthia Chew for assisting with running some of the statistical analyses, the two JMIR peer reviewers for their extremely valuable and constructive feedback, all authors of JMIR who continue to submit amazing work to JMIR, and the JMIR followers and other twitter users who help disseminating their work.

Conflicts of Interest

The author is editor and publisher of JMIR and is a shareholder of JMIR Publications Inc., which owns and publishes JMIR. He does not currently take any salary from JMIR Publications Inc., but his wife does (and complains it isn't enough). JMIR Publications Inc. also owns the domains twimpact.org, twimpactfactor.org and twimpactfactor.com with the possible goal to create services to calculate and track twimpact and twindex metrics for publications and publishers, and may or may not directly or indirectly profit from these services.

Multimedia Appendix 1

Dataset of 4208 tweets mentioning JMIR articles between July 24, 2008 and November 20, 2011. Real-time, daily updated summary article-level metrics for all JMIR articles including citation and tweetation counts, twimpact factors, twindex etc. can be downloaded from http://www.jmir.org/stats/articles?format=csv. Researchers interested in using and further analyzing these data are encouraged to contact the author.

[XLSX File (Microsoft Excel File), 540KB - jmir_v13i4e123_app1.xlsx]

XSL•FO
RenderX

Multimedia Appendix 2

Included references.

[PDF File (Adobe PDF File), 443KB - jmir_v13i4e123_app2.pdf]

References

1. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 2005 Nov 15;102(46):16569-16572. [doi: 10.1073/pnas.0507655102] [Medline: 16275915]
2. Garfield E. The history and meaning of the journal impact factor. JAMA 2006 Jan 4;295(1):90-93. [doi: 10.1001/jama.295.1.90] [Medline: 16391221]
3. Rossner M, Van Epps H, Hill E. Show me the data. J Cell Biol 2007 Dec 17;179(6):1091-1092 [FREE Full text] [doi: 10.1083/jcb.200711140] [Medline: 18086910]
4. PLoS Editors. The impact factor game. It is time to find a better way to assess the scientific literature. PLoS Med 2006 Jun;3(6):e291 [FREE Full text] [doi: 10.1371/journal.pmed.0030291] [Medline: 16749869]
5. Smith R. Measuring the social impact of research. BMJ 2001 Sep 8;323(7312):528 [FREE Full text] [Medline: 11546684]
6. Niederkrotenthaler T, Dorner TE, Maier M. Development of a practical tool to measure the impact of publications on the society based on focus group discussions with scientists. BMC Public Health 2011;11:588 [FREE Full text] [doi: 10.1186/1471-2458-11-588] [Medline: 21787432]
7. Thelwall M. Introduction to Webometrics: Quantitative Web Research for the Social Sciences. In: Synthesis Lectures on Information Concepts, Retrieval, and Services. San Francisco, USA: Morgan and Claypool Publishers; 2009.
8. Vaughan, L. Web link counts correlate with ISI impact factors: Evidence from two disciplines. Proceedings of the American Society for Information Science and Technology 2005;39:436. [doi: 10.1002/meet.1450390148]
9. Vaughan L, Shaw D. Bibliographic and Web citations: What is the difference? Journal of the American Society for Information Science and Technology 2003;54(14):1313-1322. [doi: 10.1002/asi.10338]
10. Vaughan L, Shaw D. Web citation data for impact assessment: A comparison of four science disciplines. Journal of the American Society for Information Science and Technology 2005;56(10):1075-1087. [doi: 10.1002/asi.20199]
11. Kousha K, Thelwall M, Rezaie S. Using the Web for research evaluation: The Integrated Online Impact indicator. Journal of Informetrics 2010;4(1):124-135. [doi: 10.1016/j.joi.2009.10.003]
12. Thelwall M, Kousha K. Online presentations as a source of scientific impact?: An analysis of PowerPoint files citing academic journals. Journal of the American Society for Information Science and Technology 2008;59(5):805-815. [doi: 10.1002/asi.v59:5]
13. Kousha K, Thelwall M. Google Book Search: Citation analysis for social science and the humanities. Journal of the American Society for Information Science and Technology 2009;60(8):1537-1549. [doi: 10.1002/asi.v60:8]
14. Priem J, Hemminger BM. Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. First Monday 2010;15(7):7-5 [FREE Full text]
15. Li X, Thelwall M, Giustini D. Validating Online Reference Managers for Scholarly Impact Measurement (FP). 2011 Presented at: ISSI 2011 Conference; 4-7 July 2011; Durban.
16. Priem J, Taraborelli D, Groth P, Neylon C. altmetrics.org. alt-metrics: a manifesto URL: http://altmetrics.org/manifesto/ [accessed 2011-12-11] [WebCite Cache ID 63rSi4m5h]
17. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
18. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
19. Hubbard DW. Pulse: The New Science of Harnessing Internet Buzz to Track Threats and Opportunities. Hoboken, NJ: Wiley; 2011.
20. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248. [Medline: 17238340]
21. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
22. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
23. Asur S, Huberman BA. Predicting the future with social media.: HP Labs URL: http://www.hpl.hp.com/research/scl/papers/socialmedia/socialmedia.pdf [accessed 2011-12-04] [WebCite Cache ID 63h6XZ69B]
24. Priem J, Costello KL. How and why scholars cite on Twitter. Proceedings of the American Society for Information Science and Technology 2010;47:1-4. [doi: 10.1002/meet.14504701201]
25. Letierce J, Passant A, Decker S, Breslin JG. Understanding how Twitter is used to spread scientific messages. 2010 Presented at: Proceedings of the Web Science Conference (WebSci10): Extending the Frontiers of Society On-Line; 2010; Raleigh, NC, USA URL: http://journal.webscience.org/314/2/websci10_submission_79.pdf

XSL•FO
RenderX

26. Weller K, Dröge E, Puschmann C. Citation analysis in Twitter: Approaches for defining and measuring information flows within tweets during scientific conferences. 2011. 1st Workshop on Making Sense of Microposts URL: http://ceur-ws.org/Vol-718/paper_04.pdf [accessed 2011-12-12] [WebCite Cache ID 63t8eIUMG]
27. Ross C, Terras M, Warwick C, Welsh A. UCL, self-archived draft version. 2011. Enabled backchannel: Conference Twitter use by digital humanists URL: http://www.ucl.ac.uk/infostudies/claire-ross/Digitally_Enabled_Backchannel.pdf [accessed 2011-12-12] [WebCite Cache ID 63s9cmI4k]
28. Weller K, Puschmann C. Twitter for scientific communication: How can citations/references be identified and measured? In: Proceedings of the Poster Session. 2011 Presented at: Web Science Conference; 2011; Koblenz, Germany URL: http://journal.webscience.org/500/1/153_paper.pdf
29. Priem J, Costello K, Dzuba T. Prevalence and use of Twitter among scholars. In: self-archived Poster. 2011 Presented at: Metrics: Symposium on Informetric and Scientometric Research; 2011; New Orleans, LA, USA URL: http://jasonpriem.org/self-archived/5uni-poster.png
30. Priem J, Piwowar H, Hemminger B. Altmetrics in the wild: An exploratory study of impact metrics based on social media. 2011 Presented at: Metrics: Symposium on Informetric and Scientometric Research; 2011; New Orleans, LA, USA.
31. Newmann MEJ. Power laws, Pareto distributions and Zipf's law. Contemporary Physics 2005;46:323-351. [doi: 10.1080/00107510500052444]
32. Kousha K, Thelwall M. Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. Scientometrics 2007;74(2):273-294. [doi: 10.1007/s11192-008-0217-x]
33. Ripberger JT. Social Science Research Network. 2010. Public Attention and Search-Based Infoveillance URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1539137 [accessed 2010-06-17] [WebCite Cache ID 5qYaLglnG]
34. Ripberger JT. Capturing Curiosity: Using Internet Search Trends to Measure Public Attentiveness. Policy Studies Journal 2011;39(2):239-259. [doi: 10.1111/j.1541-0072.2011.00406.x]
35. Eysenbach G. The open access advantage. J Med Internet Res 2006;8(2):e8 [FREE Full text] [doi: 10.2196/jmir.8.2.e8] [Medline: 16867971]

Abbreviations

API: application programming interface
ctw7/75: collective twimpact factor for the 75th percentile of cumulative tweets within first 7 days of publication for all articles in a set of articles
DOI: digital object identifier
fn: false negative
fp: false positive
JMIR: Journal of Medical Internet Research
PLoS: Public Library of Science
THLn: tweeted half-life
tn: true negative
tp: true positive
tw*n*: cumulative number of tweets within *n* days of publication of an article, with day 0 being the publication date; tw7 is also called the twimpact factor

Edited by A Federer; submitted 22.11.11; peer-reviewed by M Thelwall, J Priem; comments to author 11.12.11; revised version received 12.12.11; accepted 12.12.11; published 16.12.11.

Please cite as:
Eysenbach G
Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact
J Med Internet Res 2011;13(4):e123
URL: http://www.jmir.org/2011/4/e123/
doi:10.2196/jmir.2012
PMID:22173204

©Gunther Eysenbach. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 16.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Editorial

CONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions

Gunther Eysenbach[1,2,3,4], MD, MPH, FACMI; CONSORT-EHEALTH Group[4]

[1]University Health Network, Centre for Global eHealth Innovation & Techna Institute, Toronto, ON, Canada

[2]Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

[3]JMIR Publications Inc., Toronto, ON, Canada

[4]see Acknowledgments for contributors

Corresponding Author:
Gunther Eysenbach, MD, MPH, FACMI
University Health Network
Centre for Global eHealth Innovation & Techna Institute
190 Elizabeth St
Toronto, ON, M4L3Y7
Canada
Phone: 1 416 7866970
Fax: 1 416 340 3595
Email: geysenba@uhnres.utoronto.ca

Abstract

Background: Web-based and mobile health interventions (also called “Internet interventions” or "eHealth/mHealth interventions") are tools or treatments, typically behaviorally based, that are operationalized and transformed for delivery via the Internet or mobile platforms. These include electronic tools for patients, informal caregivers, healthy consumers, and health care providers. The Consolidated Standards of Reporting Trials (CONSORT) statement was developed to improve the suboptimal reporting of randomized controlled trials (RCTs). While the CONSORT statement can be applied to provide broad guidance on how eHealth and mHealth trials should be reported, RCTs of web-based interventions pose very specific issues and challenges, in particular related to reporting sufficient details of the intervention to allow replication and theory-building.

Objective: To develop a checklist, dubbed CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth), as an extension of the CONSORT statement that provides guidance for authors of eHealth and mHealth interventions.

Methods: A literature review was conducted, followed by a survey among eHealth experts and a workshop.

Results: A checklist instrument was constructed as an extension of the CONSORT statement. The instrument has been adopted by the *Journal of Medical Internet Research* (JMIR) and authors of eHealth RCTs are required to submit an electronic checklist explaining how they addressed each subitem.

Conclusions: CONSORT-EHEALTH has the potential to improve reporting and provides a basis for evaluating the validity and applicability of eHealth trials. Subitems describing how the intervention should be reported can also be used for non-RCT evaluation reports. As part of the development process, an evaluation component is essential; therefore, feedback from authors will be solicited, and a before-after study will evaluate whether reporting has been improved.

(J Med Internet Res 2011;13(4):e126) doi:10.2196/jmir.1923

KEYWORDS

evaluation; Internet; mobile health; reporting standards; publishing standards; guidelines; quality control; randomized controlled trials as topic; medical informatics

XSL•FO
RenderX

Introduction

Introducing CONSORT-EHEALTH

The current issue of the *Journal of Medical Internet Research* (JMIR) 4/2011 is the first issue where we asked JMIR authors describing randomized trials to report their trials in accordance with a new instrument designed to improve the quality of reporting of eHealth and mHealth trials, dubbed CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth). While completing the checklist is only mandatory for authors of randomized controlled trials (RCTs), the checklist may also be useful for researchers employing other evaluation methods. Beyond web-based and mobile applications, the checklist presented here is probably also applicable for a wide range of other medical informatics and technology applications. This editorial provides a short introduction on the rationale and background of this ongoing initiative to improve the quality of research and to improve knowledge translation in this field.

Web-based and Mobile Health Interventions

Web-based health interventions (also called "Internet interventions" or "eHealth interventions") are, for the purpose of this paper, "treatments, typically behaviorally based, that are operationalized and transformed for delivery via the Internet" [1]. With mobile devices being an increasingly important access point for Internet-based or otherwise networked electronic interventions, this definition includes interventions that are delivered through mobile devices or the new generation of tablet computers (eg, the iPad). Examples are behavior change interventions that help people quit smoking or lose weight, or mental health applications to address depression, anxieties, or other important health problems. An increasingly important area is the management of chronic diseases such as diabetes using Internet-based or mobile disease management programs, as well as patient-accessible personal health records, and tailored educational programs for patients. The term "treatment" should be understood in its broadest sense, and includes, for example, management tools, electronic tools that improve the communication (eg, between patient or health professionals) or systems that provide diagnostic or prognostic information or aid in the triage of patients.

Web-based and mobile interventions are increasingly important instruments in the toolkit of public health professionals and researchers [1-3]. The web-based delivery mode makes it relatively easy to enroll and track a large number of participants in longitudinal studies, including RCTs, to test the effectiveness of specific program components, or to evaluate the effectiveness of the program as a whole. The ease of enrollment comes, however, at a cost: compared to face-to-face trials, researchers in eHealth trials have less control over the participants, and Internet-based trials pose some other specific problems, related to execution and reporting of the trial [4].

While this is a young field, with less than a dozen web-based RCTs published before 2002 [4, 5], the number of reports evaluating web-based interventions in the medical literature is increasing rapidly. In October 2010, a scan of literature indexed in PubMed with the publication type "randomized trial" and major medical subject headings (MeSH) term "Internet", elicited 582 published randomized trials. This does not take into account evaluations of mobile networked applications (which may not be indexed with the "Internet" keyword), or studies with nonrandomized longitudinal designs.

In addition to the rapidly growing area of Internet interventions, RCTs evaluating non-Internet interventions are also using elements of web-based trials, for example, Web-based recruitment or Web-based data collection [6].

While JMIR is the leading journal in this field (in terms of both impact and number of articles published in this field), these trials are scattered across a wide variety of journals: in the October 2010 scan, 263 different journals were identified which have published at least one "eHealth RCT". While JMIR was the journal which had most trials published, the distribution has a very long tail, with relevant articles scattered in a wide range of other journals (see Table 1). As a consequence, reporting standards and the level of detail provided in these publications vary widely, hampering progress in this area, and impeding knowledge translation. While at JMIR we are requiring authors to submit the CONSORT (Consolidated Standards of Reporting Trials) checklist [7-11] and use additional checklists for some aspects of these trials (eg, Checklist for Reporting Results of Internet E-Surveys [CHERRIES] [12]), internationally developed and adopted reporting guidelines specifically for eHealth and mHealth trials are lacking.

The CONSORT statement was developed to improve the suboptimal reporting of RCTs [9]. While the CONSORT statement [9] and some published extensions [7,8,10,11] as well as other guidelines for other study designs and domains can be applied to provide broad guidance on how such evaluations should be reported, RCTs of web-based interventions pose very specific issues and challenges [4, 13], which we intended to amalgamate and elaborate in the form of a eHealth-specific CONSORT extension guideline and checklist, called CONSORT-EHEALTH.

XSL•FO
RenderX

Table 1. Ranking of journals which have published at least 4 randomized trials of web-based applications (indexed with "Internet" as major MeSH heading and publication type = RCT) [from a list of a total of 582 trial publications], as of October 2010 (journal titles as per PubMed)

Journal name	number of Internet RCTs
Journal of medical Internet research (JMIR)	37
Preventive medicine	12
Journal of consulting and clinical psychology	11
Nicotine & tobacco research: official journal of the Society for Research on Nicotine and Tobacco	10
Diabetes care	9
Health education research	9
Behaviour research and therapy	9
Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society	9
Academic medicine: journal of the Association of American Medical Colleges	8
Journal of health communication	8
Cognitive behavior therapy	8
The Australian and New Zealand journal of psychiatry	7
BMC psychiatry	7
Studies in health technology and informatics	7
Annals of behavioral medicine: a publication of the Society of Behavioral Medicine	7
Pediatrics	6
Patient education and counseling	6
Addiction (Abingdon, England)	6
Journal of substance abuse treatment	6
American journal of preventive medicine	6
BMC medical education	5
Obesity (Silver Spring, Md.)	5
Behavior research methods, instruments, & computers: a journal of the Psychonomic Society, Inc	5
Archives of internal medicine	5
Addictive behaviors	5
Journal of nutrition education and behavior	5
Proceedings / AMIA ... Annual Symposium. AMIA Symposium	5
The Journal of adolescent health: official publication of the Society for Adolescent Medicine	5
BMC public health	5
AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium	5
Journal of general internal medicine	5
Trials	5
JAMA: the journal of the American Medical Association	4
The British journal of psychiatry: the journal of mental science	4
Journal of pediatric psychology	4
Journal of clinical nursing	4
Health psychology: official journal of the Division of Health Psychology, American Psychological Association	4
Computers, informatics, nursing: CIN	4
Depression and anxiety	4
Journal of telemedicine and telecare	4

XSL•FO
RenderX

Methods

To develop a guideline for reports of web-based interventions we broadly followed the standard methodology developed by the CONSORT group, reported in detail elsewhere [14]. We started the work on CONSORT-EHEALTH in October 2010 with writing a grant proposal requesting funding for a consensus workshop from the Canadian Institutes of Health Research (CIHR). Unfortunately, this funding request was turned down (with some rather odd explanations, such as "[it is unclear] why journal editors [private sector] need funding to complete this project.") Without funding, our initial plan to use a 3-phase process of premeeting item generation, a meeting with invited stakeholders, and postmeeting consolidation, had to be modified, with only a very short face-to-face workshop in the context of a scientific meeting, and the bulk of the work being conducted through online consultations.

The core international group of CONSORT-EHEALTH contributors included researchers, funders, consumers, journal editors, and industry, listed under acknowledgments. This is (and remains) an open and dynamic group.

In the premeeting item-generation process, we used the current CONSORT guideline items as a framework, and generated additional items and subitems through literature searches, extracting reported items from published RCTs as well as relevant guidelines. We had access to the referee reports of JMIR, which helped us to evaluate which items are frequently pointed out by reviewers as "missing" in the original submissions of the authors. Additional input came from a face-to-face session hosted by the International Society for Research on Internet Interventions (ISRII), in Sydney, Australia on April 6-8, 2011.

A preliminary version of the CONSORT-EHEALTH checklist (V1.5) was published in April 2010. In a web-based Delphi process we gathered some data on the importance of the items [14]. The initial questionnaire with the list of items is shown in Multimedia Appendix 1. Participants were asked to suggest missing items (under each CONSORT subheading), and to rate each proposed subitem on a scale of 1-5 (where 1 was "subitem not at all important" and 5 was "essential").

We kept items as "essential" in the CONSORT-EHEALTH when at least 50% of respondents rated an item as "5-essential". We downgraded items as "highly recommended" when at least 50% of respondents rated an item as 4 or 5 (but less than 50% said it is "essential"). We eliminated items when less than 50% of respondents answered 4 or 5.

Results

CONSORT-EHEALTH

The key references identified and used for item generation are the original CONSORT items–in particular including the extension for nonpharmacologic treatments [7]–as well as an early paper by Eysenbach on eHealth-specific RCT issues [4], a paper on the relevance of CONSORT reporting criteria for eHealth trials by Baker and colleagues [13], and a preprint of a recent paper by Proudfoot and colleagues [15]. An existing guideline for evaluation studies in health informatics [16] proved to be too broad to be of much use. Systematic reviews in this area [2, 3, 17, 18] provided further useful frameworks for identifying items that should be reported.

Based on these references, we published an initial instrument (CONSORT-EHEALTH V1.5) with 53 additional subitems, either explaining or enhancing the original 25 CONSORT items. EHEALTH-specific subitems were indicated with Roman numerals (eg, CONSORT item 2a had two additional subitems numbered 2a-i and 2a-ii). We added 2 items to the original 25-item CONSORT (item X26 on ethics, and item X27 on conflict of interest disclosure), which are not part of the original CONSORT checklist and did not fit under any existing item. This instrument was published in March 2011, and JMIR readers and eHealth stakeholders were asked to comment on the instrument and to rate the importance of each subitem.

Between April 4 and June 17 2011, we received 55 responses to CONSORT V1.5 (first Delphi round). Multimedia Appendix 2 and Multimedia Appendix 3 show the responses.

Most users agreed with all subitems. Only 1 subitem was eliminated after the first Delphi round ("Report how institutional affiliations are displayed to potential participants [on eHealth media], as affiliations with prestigious hospitals or universities may affect volunteer rates, use, and reactions with regard to an intervention").

The resulting current iteration of CONSORT-EHEALTH V1.6.1 (which is currently in use at JMIR) has 17 subitems that are deemed "essential", and 35 subitems that are deemed "highly recommended" (Multimedia Appendix 4). The checklist (V1.6.1) was published on the JMIR website on August 25, 2011 and is currently being pilot-tested with the help of JMIR authors, who are asked to submit an electronic version of the checklist via an online questionnaire when they submit reports of an RCT (Multimedia Appendix 5). In this questionnaire, authors of RCTs are required to quote passages of their manuscript corresponding to each item, or to quickly explain why they are not applicable. They are also asked to (on a voluntary basis) rate the importance of the items for their trial. We expect the next iteration of CONSORT-EHEALTH (V2.0) to be published early 2013, which will take into account the feedback received from actual RCT authors (who may also be acknowledged as part of the CONSORT-EHEALTH revision group). This way we ensure that the next iteration of CONSORT-EHEALTH is informed by the experiences of actual users.

Specific Reporting Issues of Ehealth Trials and Examples for Guideline Items

In the following we provide some examples of items and subitems that are part of the guideline.

For example, in the interest of reproducibility and comparability (in the research setting) as well as for implementation and dissemination (in practice settings), a detailed description and documentation of the intervention is required. The CONSORT checklist contains only a single item related to the intervention (item 5: "Describe the interventions for each group with sufficient details to allow replication, including how and when they were actually administered"). This may be sufficient for

XSL•FO
RenderX

drug trials, where the simple mentioning of the drug name and its administration mode and dosage is sufficient to characterize the intervention, but not sufficient for eHealth or mHealth interventions, which are complex interventions requiring more details so that others can replicate the intervention. Thus, we created a detailed subchecklist as an extension to the CONSORT item 5, listing required and desired reporting elements characterizing the functional components and other important features of the web intervention (Table 2).

Two of these subitems (5-v and 5-vi) speak to the problem of digital preservation of the intervention, which is another unique aspect of eHealth or mHealth trials. For scientific hypotheses and findings to be confirmed or disproved by other researchers, key elements should be available to other researchers, ideally as open source code, or at least be theoretically "reproducible" by disclosing algorithms, pathways of participants through the application, etc., or at a minimum by providing screenshots or archiving the interfaces in a webarchive (such as the Internet Archive or WebCitation.org). The issue of open source and complete transparency of the algorithms appeared to be somewhat controversial among respondents, as some eHealth applications may have commercial use and some respondents were concerned about publicizing proprietary algorithms. While at JMIR we highly encourage the publication of open source code alongside the publication, there does not currently seem to be a consensus to make this a universal requirement across journals. However, even if the code is not made accessible as open source, the report must contain sufficient details and preferably screenshots to allow others to replicate or disprove the key findings – otherwise it cannot be considered scientific research and should be published in a trade journal rather than a peer-reviewed scholarly journal.

Apart from expanding the intervention description item 5, there were other important additions and expansions of the current CONSORT items.

A number of guideline subitems (6a-ii, 12a-i, 13b-i, 17-i; see Multimedia Appendix 4) are related to the important issue of attrition (non-use) and use (engagement, "dose", adherence) of the intervention [19]. As participants in web-based evaluations usually have full control over whether or not they use the intervention, and how often and how long they engage with the application, real-world evaluations of web-based interventions and interpretations of reports on their effectiveness (or lack thereof) are often complicated by the fact that a substantial proportion of participants may have dropped out of the trial (non-use or loss-to-follow-up attrition) [19]. While nonadherence may be a problem in drug trials too, the attrition rates in Internet-based trials are by far higher than in trials with a face-to-face component. As effectiveness as measured in these trials is a function of (and dependent on) participants actually using the intervention, researchers should measure and report metrics of use (adherence) and/or non-use (attrition), which can be measured using a variety of metrics such as number of logins and average session time. However, even these seemingly straightforward metrics require additional explanations, for example, if researchers report an average session time, this may be skewed by some participants never logging out; therefore, additional information such as the timeout policy should be provided (eg, automatic logout after 15 minutes of inactivity) in order to enable accurate interpretation and across-trial comparisons. In subitem 6a-ii (an expansion of CONSORT item 6 "outcomes") we suggest that researchers explain how use and engagement was measured and defined, in addition to describing how the primary health outcomes were measured, and in subitem 17-i (an expansion of CONSORT item 17 "outcomes and estimation") we ask that use and usage outcomes should be reported. In subitem 12a-i (an expansion of CONSORT item 12 "statistical methods") we specifically ask how missing values due to attrition were treated statistically [20]. In addition to the traditional CONSORT flow diagram we also highly encourage the provision of an attrition diagram (CONSORT-EHEALTH item 13b-i) in the results sections, illustrating the login behavior of participants in all groups over time as a survival curve [19].

The comprehensive description of web-based recruitment strategies and data collection methods are other areas where we identified the need for guideline items. Our previously published CHERRIES guideline for reporting web-based surveys [12] may provide additional guidance and may be seen as a supplement to subitem 6a-i, which deals with the common case where outcomes were collected through online questionnaires.

There is a regrettable trend to split reports of randomized trials into "least publishable units", for example, to publish one paper with the results of the primary RCT outcomes, another paper with usage results, and another paper with a qualitative analysis of participant feedback. At JMIR, we have a strict policy against "salami publication", a practice that limits the ability of the reader to interpret the overall findings, and will consider such multipart papers only in exceptional circumstances, and preferably when the reports are submitted together and published in the same journal issue. An in-depth qualitative evaluation may justify a separate paper, but a few CONSORT-EHEALTH items (6a-iii and 19-ii) remind authors that some qualitative analysis should be part of any eHealth evaluation report, in particular if nonuse of the application or potential harmful effects were observed, which should shift the focus of the report to the question *why* these results occurred.

Finally, in order to enhance retrievability (findability) of these kinds of studies in PubMed and other bibliographic databases, we also suggest preferred terms to be used in article titles and abstracts (e.g. "web-based intervention" or "mobile intervention"). These recommendations are based on an analysis of the prevalence of terms used in current studies.

XSL•FO
RenderX

Table 2. Subitems expanding CONSORT item 5 (description of intervention)

	Subitem	Importance
i)	**Mention names, credential, affiliations of the developers, sponsors, and owners** [15] (if authors/evaluators are owners or developer of the software, this needs to be declared in a "Conflict of interest" section).	Highly Recommended
ii)	**Describe the history/development process** of the application and previous formative evaluations (e.g., focus groups, usability testing), as these will have an impact on adoption/use rates and help with interpreting results.	Highly Recommended
iii)	**Revisions and updating**. Clearly mention the date and/or version number of the application/intervention (and comparator, if applicable) evaluated, or describe whether the intervention underwent major changes during the evaluation process, or whether the development and/or content was "frozen" during the trial. Describe dynamic components such as news feeds or changing content which may have an impact on the replicability of the intervention (for unexpected events see item 3b).	Highly Recommended
iv)	Provide information on **quality assurance methods** to ensure accuracy and quality of information provided [13], if applicable.	Highly Recommended
v)	**Ensure replicability by publishing the source code (preferably as open source), and/or providing screenshots/screen-capture video, and/or providing flowcharts of the algorithms used.** Replicability (i.e., other researchers should in principle be able to replicate the study) is a hallmark of scientific reporting.	Highly Recommended
vi)	**Digital preservation**: Provide the URL of the application, but as the intervention is likely to change or disappear over the course of the years, also make sure the intervention is archived (Internet Archive, webcitation.org, and/or publishing the source code or screenshots/videos alongside the article). As pages behind login screens cannot be archived, consider creating demo pages which are accessible without login.	Highly Recommended
vii)	**Access**: Describe how participants accessed the application, in what setting/context, if they had to pay (or were paid) or not, whether they had to be a member of specific group. If known, describe how participants obtained "access to the platform and Internet" [13]. To ensure access for editors/reviewers/readers, consider providing a "backdoor" login account or demo mode for reviewers/readers to explore the application (also important for archiving purposes, see vi).	Essential
viii)	Describe **mode of delivery, features/functionalities/components of the intervention and comparator, and the theoretical framework** [6] used to design them (instructional strategy [13], behavior change techniques, persuasive features, etc., see e.g., [17, 18] for terminology). This includes an in-depth description of the content (including where it is coming from and who developed it) [13], "whether [and how] it is tailored to individual circumstances and allows users to track their progress and receive feedback" [15]. This also includes a description of communication delivery channels and – if computer-mediated communication is a component – whether communication was synchronous or asynchronous [15]. It also includes information on presentation strategies [13], including page design principles, average amount of text on pages, presence of hyperlinks to other resources etc. [13].	Essential
ix)	**Describe use parameters** (e.g., intended "doses" and optimal timing for use) [13]. Clarify what instructions or recommendations were given to the user, for example, regarding timing, frequency, heaviness of use [13], if any, or was the intervention used ad libitum.	Highly Recommended
x)	**Clarify the level of human involvement** (care providers or health professionals, also technical assistance) in the e-intervention or as co-intervention. Detail number and expertise of professionals involved, if any, as well as "type of assistance offered, the timing and frequency of the support, how it is initiated, and the medium by which the assistance is delivered" [15]. It may be necessary to distinguish between the level of human involvement required for the trial, and the level of human involvement required for a routine application outside of an RCT setting (discuss under item 21 – generalizability).	Highly Recommended
xi)	**Report any prompts/reminders used:** Clarify if there were prompts (letters, emails, phone calls, SMS) to use the application, what triggered them, frequency, etc. [13]. It may be necessary to distinguish between the level of prompts/reminders required for the trial, and the level of prompts/reminders for a routine application outside of an RCT setting (discuss under item 21 – generalizability).	Essential
xii)	**Describe any co-interventions (including training/support)**: Clearly state any "interventions that are provided in addition to the targeted eHealth intervention" [13], as eHealth intervention may not be designed as stand-alone intervention. This includes training sessions and support [13]. It may be necessary to distinguish between the level of training required for the trial, and the level of training for a routine application outside of an RCT setting (discuss under item 21 – generalizability).	Essential

Discussion

We hypothesize that publication of the guideline in August 2010 will have a significant impact on the quality of reports of web-based intervention evaluations, which will in turn enable better systematic reviews and facilitate knowledge translation. The guideline will hopefully also be a useful starting point and framework for discussions around the quality of eHealth trials, how such trials should actually be conducted, which items should be reported in protocols, grant proposals and trial registries, and how trials should be classified and synthesized in systematic reviews.

XSL•FO
RenderX

Elements of the guideline may be useful for researchers of other disciplines who use web-based recruitment or data collection methods, even if it is not an Internet- or mobile intervention which is being evaluated.

Many elements of the guideline (particularly the section describing subitems of the intervention) are applicable not only to randomized trials, but any kind of evaluation report.

While the *Journal of Medical Internet Research* is the first journal to adopt these guidelines, we hope that other journals and organizations endorse and adopt the guidelines. Authors are encouraged to report their research (and research protocols) in accordance with CONSORT-EHEALTH, regardless of the ultimate publication venue. Authors preparing their reports in accordance with CONSORT-EHEALTH are encouraged to cite the current guidelines (this paper), in order to facilitate further dissemination and uptake of best practices for reporting.

The current checklist is only the first step and the guideline will be very much a living document in an iterative and ongoing development process. As technology is changing constantly and rapidly, and reporting of eHealth and mHealth interventions is determined by what is technologically possible, the checklist will need to be updated much more frequently than other guidelines dealing with more "static" interventions, such as acupuncture (STRICTA) [21].

As part of the iterative development process, an ongoing evaluation component is essential; otherwise, asking authors, journals and editors to use and endorse the guidance is not warranted.

To provide a body of evidence to support usage of the guideline we intend to evaluate, elaborate on, and further develop the CONSORT-EHEALTH checklist, by

- a pilot implementation at the *Journal of Medical Internet Research* which involves collecting data from RCT authors (this pilot started with the current issue 4/2011) (see Multimedia Appendix 5 for the data collection form)
- a retrospective analysis of a random sample of web-based RCTs, published before publication of the CONSORT-EHEALTH checklist (ongoing)
development of an Explanation and Elaboration manuscript
- development of a website and an interactive toolkit
- the formation of a standing working group to lead the continued development of the guideline
- a systematic analysis of RCTs of web-based interventions published after publication of the guideline, to evaluate the impact of the checklist, and to identify shortcomings and new items
- creation of a searchable database of trials (based on the information entered by JMIR authors when filling in the CONSORT-EHEALTH checklist).

It should be stressed again that the development of CONSORT-EHEALTH is an iterative and ongoing process, which requires a broad stakeholder input, which we welcome. We will continue to try to obtain funding for this important work which in our view is essential to advance the art and science of Internet and mobile interventions.

Acknowledgments

We thank David Moher and the CONSORT team, and the International Society for Research on Internet Interventions (ISRII) leadership, in particular Lee Ritterband and Helen Christensen, for supporting this project. We also thank the participants of the ISRII workshop in Sydney, in April 2011, and those who filled in the questionnaire. The following respondents opted to have their name acknowledged:

Susan Stoner; Talaria, Inc., Seattle, WA, USA

Filip Drozd; University of Oslo, Oslo, Norway

Matthijs Blankers; Academic Medical Hospital, University of Amsterdam, Amsterdam, The Netherlands

Rik Crutzen; Maastricht University/CAPHRI, Maastricht, The Netherlands

Robert Tait; Australian National University ACT, Australia

David Mohr; Northwestern University, USA

Doug Evans; George Washington University, USA

Robin Kok; VU University Amsterdam, The Netherlands

Sorayouth Chumnanvej; Mahidol University, BKK, Thailand

Pongsak Kooprasertying; Drug Information Service, Bumrungrad International Hospital, Thailand

Stephen Anthony; The University of New South Wales, Sydney, Australia

Lucy Wanjuki Kivuti; University of Nairobi, Kenya

Karen Hambly; University of Kent, Medway, UK

XSL•FO
RenderX

J.W.M. Aarts; Radboud University Nijmegen Medical Center, The Netherlands

Judith Prins; University Medical Centre St Radboud, Nijmegen, The Netherlands

Isabel de la Torre; University of Valladolid, Valladolid, Spain

Royer Cook; ISA Associates, USA

Betty C. Jung; Connecticut, USA

Richard Burkhard; San Jose State University, CA, USA

Armando Ruiz-Baques; Autonomous University of Barcelona, Barcelona, Spain

Elia Gabarron; Autonomous University of Barcelona, Barcelona, Spain

Alireza Ahmadvand; Tehran University of Medical Sciences, Tehran, Iran

Stephen Gentles; McMaster University, Canada

Sylvia Kauer; UniMelb & MCRI, Melbourne, Australia

Ólöf Birna Kristjánsdóttir; Oslo University College, Oslo, Norway

Rajneesh Choubisa; Indian Institute of Technology, Delhi, India

Emmanuel Kusi Achampong; University of Cape Coast, Ghana

Daniel Arvidsson; Center for Primary Health Care Research, Malmö, Sweden

Marie-Pierre Gagnon; Université Laval, Québec, Canada

Mahmut Gurgan; Bezmialem Vakif University, Istanbul, Turkey

Steven Daviss; University of Maryland, USA

Erin Mathieu; University of Sydney, Sydney, Australia

Michel Labrecque; Université Laval, Québec, Canada

Hansen Gottlieb Anders; University of Southern Denmark

Tuomas Lehto; University of Oulu, Oulu, Finland

Eleni Kaldoudi; Democritus University of Thrace, Alexandroupoli, Greece

Atle Klovning; University of Oslo, Oslo, Norway

Ming-Yuan Chih; University of Wisconsin Madison, Madison, WI, USA

Konstantinos Antypas; NST, Tromsø, Norway

Hilde Eide; Buskerud University College, Drammen, Norway

Terry Lewin; University of Newcastle, Newcastle, Australia

Fung Lin Elean Leung; Chinese University of Hong Kong, Hong Kong, China

Jessica Ancker; Weill Cornell Medical College, USA

Juanita Fernando; Monash University, Australia

Patrick M Archambault; Université Laval, Canada

Rodrigo Dias; Department and Institute of Psychiatry, São Paulo University Medical School, Brazil

Annie Lau; University of New South Wales, Sydney, Australia

Peter Beck; Joanneum Research, Graz, Austria

Evan Mayo-Wilson; University of Oxford, UK

Brigitte Bewick; University of Leeds, Leeds, UK

Conflicts of Interest

The author is editor and publisher of JMIR.

XSL•FO
RenderX

Multimedia Appendix 1

Online questionnaire of the initial CONSORT-EHEALTH instrument V1.5.

[PDF File (Adobe PDF File), 185KB - jmir_v13i4e126_app1.pdf]

Multimedia Appendix 2

Summary of responses from the online questionnaire of CONSORT-EHEALTH V1.5 (note that narrative responses are excerpts only).

[PDF File (Adobe PDF File), 502KB - jmir_v13i4e126_app2.pdf]

Multimedia Appendix 3

Database with all responses received between April 4 and June 17, 2011, in response to the CONSORT-EHEALTH V1.5 questionnaire (email addresses and certain names removed).

[XLS File (Microsoft Excel File), 233KB - jmir_v13i4e126_app3.xls]

Multimedia Appendix 4

Current CONSORT-EHEALTH V1.6.1 checklist.

[PDF File (Adobe PDF File), 359KB - jmir_v13i4e126_app4.pdf]

Multimedia Appendix 5

Submission form for JMIR (based on CONSORT-EHEALTH V1.6), available at [http://tinyurl.com/consort-ehealth-v1-6] .

[PDF File (Adobe PDF File), 211KB - jmir_v13i4e126_app5.pdf]

References

1. Ritterband LM, Andersson G, Christensen HM, Carlbring P, Cuijpers P. Directions for the International Society for Research on Internet Interventions (ISRII). J Med Internet Res 2006;8(3):e23 [FREE Full text] [doi: 10.2196/jmir.8.3.e23] [Medline: 17032639]
2. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
3. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
4. Eysenbach G. Issues in evaluating health websites in an Internet-based randomized controlled trial. J Med Internet Res 2002 Dec;4(3):E17 [FREE Full text] [doi: 10.2196/jmir.4.3.e17] [Medline: 12554548]
5. Bessell TL, McDonald S, Silagy CA, Anderson JN, Hiller JE, Sansom LN. Do Internet interventions for consumers cause more harm than good? A systematic review. Health Expect 2002 Mar;5(1):28-37. [Medline: 11906539]
6. El Emam K, Jonker E, Sampson M, Krleza-Jerić K, Neisa A. The use of electronic data capture tools in clinical trials: Web-survey of 259 Canadian trials. J Med Internet Res 2009;11(1):e8 [FREE Full text] [doi: 10.2196/jmir.1120] [Medline: 19275984]
7. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008 Feb 19;148(4):295-309 [FREE Full text] [Medline: 18283207]
8. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, CONSORT group, Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008;337:a2390 [FREE Full text] [Medline: 19001484]
9. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869. [Medline: 20332511]
10. Campbell MK, Elbourne DR, Altman DG, CONSORT group. CONSORT statement: extension to cluster randomised trials. BMJ 2004 Mar 20;328(7441):702-708 [FREE Full text] [doi: 10.1136/bmj.328.7441.702] [Medline: 15031246]
11. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA 2006 Mar 8;295(10):1152-1160. [doi: 10.1001/jama.295.10.1152] [Medline: 16522836]

XSL•FO
RenderX

12. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
13. Baker TB, Gustafson DH, Shaw B, Hawkins R, Pingree S, Roberts L, et al. Relevance of CONSORT reporting criteria for research on eHealth interventions. Patient Educ Couns 2010 Dec;81 Suppl:S77-S86. [doi: 10.1016/j.pec.2010.07.040] [Medline: 20843621]
14. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med 2010 Feb;7(2):e1000217 [FREE Full text] [doi: 10.1371/journal.pmed.1000217] [Medline: 20169112]
15. Proudfoot J, Klein B, Barak A, Carlbring P, Cuijpers P, Lange A, et al. Establishing Guidelines for Executing and Reporting Internet Intervention Research. Cognitive Behaviour Therapy 2011;40:82-97. [doi: 10.1080/16506073.2011.573807]
16. Talmon J, Ammenwerth E, Brender J, de Keizer N, Nykänen P, Rigby M. STARE-HI--Statement on reporting of evaluation studies in Health Informatics. Int J Med Inform 2009 Jan;78(1):1-9. [doi: 10.1016/j.ijmedinf.2008.09.002] [Medline: 18930696]
17. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
18. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
19. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
20. Blankers M, Koeter MWJ, Schippers GM. Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers. J Med Internet Res 2010;12(5):e54 [FREE Full text] [doi: 10.2196/jmir.1448] [Medline: 21169167]
21. MacPherson H, Altman DG, Hammerschlag R, Youping L, Taixiang W, White A, STRICTA Revision Group. Revised STandards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): extending the CONSORT statement. J Altern Complement Med 2010 Oct;16(10):ST1-S14. [doi: 10.1089/acm.2010.1610] [Medline: 20954957]

Edited by G Eysenbach; submitted 29.08.11; this is a non–peer-reviewed article;accepted 31.12.11; published 31.12.11.

Please cite as:
Eysenbach G, CONSORT-EHEALTH Group
CONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions
J Med Internet Res 2011;13(4):e126
URL: http://www.jmir.org/2011/4/e126/
doi:10.2196/jmir.1923
PMID:22209829

©Gunther Eysenbach, CONSORT-EHEALTH Group. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Guest Editorial

WikiBuild: A New Application to Support Patient and Health Care Professional Involvement in the Development of Patient Support Tools

Patrick Michel Archambault[1,2,3,4], MD, MSc, FRCPC

[1]Centre de santé et de services sociaux Alphonse-Desjardins (Centre hospitalier affilié universitaire de Lévis), Lévis, QC, Canada
[2]Département de médecine familiale et de médecine d'urgence, Faculté de médecine, Université Laval, Québec, QC, Canada
[3]Traumatologie – urgence – soins intensifs, Centre de recherche FRSQ du CHA universitaire de Québec, Québec, QC, Canada
[4]Division de soins intensifs, Département d'anesthésiologie, Université Laval, Québec, QC, Canada

Corresponding Author:
Patrick Michel Archambault, MD, MSc, FRCPC
Centre de santé et de services sociaux Alphonse-Desjardins (Centre hospitalier affilié universitaire de Lévis)
143 rue Wolfe
Lévis, QC, G6V 3Z1
Canada
Phone: 1 418 835 7121 ext 3905
Fax: 1 418 835 7276
Email: patrick.m.archambault@gmail.com

Related Article:

Comment on: Gupta S, Wan FT, Newton D, Bhattacharyya OK, Chignell MH, Straus SE. WikiBuild: A New Online Collaboration Process For Multistakeholder Tool Development and Consensus Building. J Med Internet Res. 2011;13(4) p. e108 http://www.jmir.org/2011/4/e108/

Abstract

Active patient and public involvement as partners in their own health care and in the development of health services is key to achieving a health care system that is responsive to patients' needs and values. It promotes better use of the health care system, and improves health outcomes, quality of life and patient satisfaction. By involving patients and health care professionals as partners in the creation and updating of patient health support tools, wikis—highly accessible, interactive vehicles of communication—have the potential to empower users to implement these support tools in daily life. Acknowledging the potential of wikis, and recognizing that they capitalize on the free and open access to information, scientists, opinion leaders and patient advocates have suggested that wikis could help decision-making constituencies improve the delivery of health care. They might also decrease its cost and improve access to knowledge within developing countries. However, little is known about the efficacy of wikis in helping to attain these goals. There is also a need to know more about the intention of patients and health care workers to use wikis, in what circumstances and what factors will influence their use of wikis. In this issue of the Journal of Medical Internet Research, Gupta et al describe how they developed and tested a new wiki-inspired application to improve asthma care. The researchers involved patients with asthma, primary care physicians, pulmonologists and certified asthma educators in the construction of an asthma action plan. Their paper—entitled "WikiBuild: a new online collaboration process for multistakeholder tool development and consensus building"—is the first description of a wiki-inspired technology built to involve patients and health care professionals in the development of a patient support tool. This innovative study has made important contributions toward how wikis could be generalized to involve multiple stakeholders in the development of other knowledge translation tools such as clinical practice guidelines or decision aids. More specifically, Gupta et al have uncovered potential action mechanisms toward increasing usage of these tools by patients and health care professionals. These are decreasing hierarchical influences, increasing usability and adapting a tool to local context. More research is now needed to determine if the use of the resulting wiki-developed plan will actually be higher than a plan developed using other methods. Furthermore, there is also a need to assess the intention of participants to continue using wiki-based processes on an ongoing basis. It is in this dynamic and continuous retroaction loop that the support tool users—both patients and health care professionals—can adapt and improve the product after its real-life shortcomings are revealed and as new evidence becomes available. As such, a wiki would be more than a simple

XSL•FO
RenderX

patient support development tool, but could also become a dynamic and interactive repository and delivery tool that would facilitate ongoing and sustainable patient and professional engagement.

(J Med Internet Res 2011;13(4):e114) doi:10.2196/jmir.1961

KEYWORDS

Medical informatics; patient-centered care; wikis; collaborative writing applications; knowledge translation; patient and public involvement

Active patient and public involvement as partners in their own health care and in the development of health services is key to achieving a health care system that is responsive to patients' needs and values [1-3]. It promotes better use of the health care system and improves health outcomes, quality of life, and patient satisfaction [4]. By involving patients and health care professionals as partners in the creation and updating of patient health support tools, wikis—highly accessible, interactive vehicles of communication—have the potential to empower users to implement these support tools in daily life [5].

Acknowledging the potential of wikis and recognizing that they capitalize on the free and open access to information, scientists, opinion leaders, and patient advocates have suggested that wikis could help decision-making constituencies improve the delivery of health care [6,7]. Wikis might also decrease the cost of health care [8] and improve access to knowledge within developing countries [6,9,10]. However, little is known about the efficacy of wikis in helping to attain these goals. There is also a need to know more about the intention of patients and health care workers to use wikis and in what circumstances, and what factors will influence their use of wikis [11]. An ongoing scoping review on the use of wikis and collaborative writing applications in health care will better identify the areas where further knowledge synthesis is needed and the areas where more primary research remains to be done [12].

In this issue of the *Journal of Medical Internet Research*, Gupta et al [13] describe how they developed and tested a new wiki-inspired application to improve asthma care. The researchers involved patients with asthma, primary care physicians, pulmonologists, and certified asthma educators in the construction of an asthma action plan. Their paper—entitled "WikiBuild: a new online collaboration process for multistakeholder tool development and consensus building"—is the first description of a wiki-inspired technology built to involve patients and health care professionals in the development of a patient support tool. The findings of this study will thus be an important addition to the cumulative evidence being synthesized in the ongoing scoping review [12].

Given the drive for more patient and public involvement in health care, finding effective ways to engage patients in decision making has become paramount [14]. For asthma patients, the use of an action plan—a document written by health care professionals to guide patients' individual self-management of worsening symptoms—has been shown to significantly reduce hospitalizations, emergency room visits, and missed work or school, and to significantly improve quality of life [15]. However, in practice, these asthma action plans are not used, and uptake has been low by clinicians and patients alike. Most existing action plans have been developed by teams consisting exclusively of medical experts who have focused on the content of action plans without addressing ease of use and visual design factors. By involving patients in the development of patient information materials, more relevant information can be included that is better adapted to the local context and that better meets the needs of end users [16].

In developing this custom-built application to enable peer-to-peer editing of the visual characteristics of an asthma action plan, the authors highlight the importance of the visual design of patient support tools. The way information is transmitted to patients greatly influences their decisions. Thomas Goetz brilliantly illustrates this fact in a popular TED Talk available on YouTube [17]. In this video, he points to seminal research showing that a drug facts box—a simple 1-page summary of relevant drug information—improves consumers' knowledge of prescription drug benefits and side effects [18]. The WikiBuild process proposes a bold new way of incorporating patients' and professionals' action plan design preferences with the intention of increasing its uptake.

Overall, the WikiBuild application surpassed the authors' expectations of usability in many aspects. Almost all the participants contributed to the development of the tool using the new wiki application. Even though participants had incentives to contribute, this very high contribution rate compares very well with editing rates within well-known wikis such as Wikipedia [19]. In the end, most participants were satisfied with the final action plan, and few participants perceived interstakeholder group hierarchies. One of the basic philosophies supporting the use of wikis for collaborative work is that authors are equal and authority is generally disregarded, since each contribution is judged by its merit and not by the degree or title of its author. Equality between individuals is one of the basic characteristics of collaboration, and research has shown that collaboration is hindered by power differences based on gender stereotypes and social status [20,21].

The main limitation acknowledged by the authors for their wiki-inspired application was that options in the wiki site were predetermined, possibly limiting user creativity. This limitation was intended to focus participants' attention on adapting the visual aspects. However, this constraint possibly limited participants' capacity to collaboratively write the action plan content, a process that could also increase its relevance and usability. Recognizing this vast potential, other scientists are exploring wikis to involve patients in collaborative content writing [22].

Notwithstanding this limitation, this innovative study has made important contributions toward how wikis could be generalized to involve multiple stakeholders in the development of other

XSL•FO
RenderX

knowledge translation tools such as clinical practice guidelines or decision aids. More specifically, Gupta and colleagues have uncovered potential action mechanisms toward increasing usage of these tools by patients and health care professionals. These are decreasing hierarchical influences, increasing usability, and adapting a tool to the local context.

More research is now needed to determine whether the resulting wiki-developed plan will actually be used more than a plan developed by other methods. Furthermore, there is also a need to assess the intention of participants to continue using wiki-based processes on an ongoing basis. It is in this dynamic and continuous retroactive loop that the support tool users—both patients and health care professionals— can adapt and improve the product after its real-life shortcomings are revealed and as new evidence becomes available. As such, a wiki would be more than a simple patient support development tool; it could also become a dynamic and interactive repository and delivery tool that would facilitate ongoing and sustainable patient and professional engagement.

In conclusion, Gupta and colleagues have shed new light on how wikis could engage patients and health care professionals in the creation and use of an asthma action plan. By doing so, they have also paved the way to further exploration of wikis for patient and health care professional involvement in the development of many other knowledge translation tools such as decision aids and clinical practice guidelines.

Acknowledgments

I would like to thank France Légaré, Canada Research Chair in the Implementation of Shared Decision Making in Primary Care, and Richard Fleet, Université Laval-Hôtel-Dieu de Lévis Emergency Medicine Research Chair, for having reviewed this manuscript. I must also thank Andrea Bilodeau for having helped prepare this manuscript and Sandra Marshall for editing it.

Conflicts of Interest

None declared

References

1. Légaré F, Boivin A, van der Weijden T, Packenham C, Tapp S, Burgers J. A knowledge synthesis of patient and public involvement in clinical practice guidelines: study protocol. Implement Sci 2009;4:30 [FREE Full text] [doi: 10.1186/1748-5908-4-30] [Medline: 19497114]
2. van der Weijden T, Légaré F, Boivin A, Burgers JS, van Veenendaal H, Stiggelbout AM, et al. How to integrate individual patient values and preferences in clinical practice guidelines? A research protocol. Implement Sci 2010;5:10 [FREE Full text] [doi: 10.1186/1748-5908-5-10] [Medline: 20205815]
3. British Medical Association. 2011 Jun. Patient and Public Involvement - A Toolkit for Doctors URL: http://www.bma.org.uk/images/ppitoolkit_tcm41-207789.pdf [accessed 2011-09-13] [WebCite Cache ID 61gNXyq2t]
4. Harkness J. Patient involvement: a vital principle for patient-centred health care. World Hosp Health Serv 2005;41(2):12-6, 40. [Medline: 16104453]
5. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
6. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]
7. Czarnecka-Kujawa K, Abdalian R, Grover SC. M1042 The quality of open access and open source Internet material in gastroenterology: is Wikipedia appropriate for knowledge transfer to patients? Gastroenterology 2008;134(4, Suppl 1):A325-A326. [doi: 10.1016/S0016-5085(08)61518-8]
8. Wright A, Bates DW, Middleton B, Hongsermeier T, Kashyap V, Thomas SM, et al. Creating and sharing clinical decision support content with Web 2.0: Issues and examples. J Biomed Inform 2009 Apr;42(2):334-346. [doi: 10.1016/j.jbi.2008.09.003] [Medline: 18935982]
9. de Silva V, Hanwella R. Why are we copyrighting science? BMJ 2010;341:c4738. [Medline: 20847026]
10. Godlee F, Pakenham-Walsh N, Ncayiyana D, Cohen B, Packer A. Can we achieve health information for all by 2015? Lancet 2004;364(9430):295-300. [doi: 10.1016/S0140-6736(04)16681-6] [Medline: 15262109]
11. Archambault PM, Légaré F, Lavoie A, Gagnon MP, Lapointe J, St-Jacques S, et al. Healthcare professionals' intentions to use wiki-based reminders to promote best practices in trauma care: a survey protocol. Implement Sci 2010;5:45 [FREE Full text] [doi: 10.1186/1748-5908-5-45] [Medline: 20540775]
12. Archambault PM, Van De Belt TH, Grajales III FJ, Eysenbach G, Aubin K, Gold I, et al. Wikis and collaborative writing applications in healthcare: a scoping review protocol. JMIR Res Protoc 2012 (Forthcoming). [doi: 10.2196/jmir.1993]
13. Gupta S, Wan FT, Newton D, Bhattacharyya OK, Chignell MH, Straus SE. WikiBuild: a new online collaboration process for multistakeholder tool development and consensus building. J Med Internet Res 2011;13(4):e108 [FREE Full text] [doi: 10.2196/jmir.1833]

XSL•FO
RenderX

14. Légaré F, Ratté S, Stacey D, Kryworuchko J, Gravel K, Graham ID, et al. Interventions for improving the adoption of shared decision making by healthcare professionals. Cochrane Database Syst Rev 2010(5):CD006732. [doi: 10.1002/14651858.CD006732.pub2] [Medline: 20464744]
15. Gibson PG, Powell H, Coughlan J, Wilson AJ, Abramson M, Haywood P, et al. Self-management education and regular practitioner review for adults with asthma. Cochrane Database Syst Rev 2003(1):CD001117. [doi: 10.1002/14651858.CD001117] [Medline: 12535399]
16. Nilsen ES, Myrhaug HT, Johansen M, Oliver S, Oxman AD. Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. Cochrane Database Syst Rev 2006;3:CD004563. [doi: 10.1002/14651858.CD004563.pub2] [Medline: 16856050]
17. Thomas Goetz. TED Talks. 2011. It's Time to Redesign Medical Data. Jan 27 URL: http://www.youtube.com/watch?v=bCGlWQnzDVE [accessed 2011-09-13] [WebCite Cache ID 61gOENpNm]
18. Schwartz LM, Woloshin S, Welch HG. Using a drug facts box to communicate drug benefits and harms: two randomized trials. Ann Intern Med 2009 Apr 21;150(8):516-527 [FREE Full text] [Medline: 19221371]
19. Priedhorsky R, Chen J, Lam SK, Panciera K, Terveen L, Riedl J. Creating, destroying, and restoring value in Wikipedia. In: Proceedings. 2007 Presented at: 2007 International ACM Conference on Supporting Group Work; November 4-7, 2007; Sanibel Island, FL, USA. [doi: 10.1145/1316624.1316663]
20. San Martín-Rodríguez L, Beaulieu MD, D'Amour D, Ferrada-Videla M. The determinants of successful collaboration: a review of theoretical and empirical studies. J Interprof Care 2005 May;19 Suppl 1:132-147. [doi: 10.1080/13561820500082677] [Medline: 16096151]
21. Légaré F, Stacey D, Gagnon S, Dunn S, Pluye P, Frosch D, et al. Validating a conceptual model for an inter-professional approach to shared decision making: a mixed methods study. J Eval Clin Pract 2011 Aug;17(4):554-564. [doi: 10.1111/j.1365-2753.2010.01515.x] [Medline: 20695950]
22. Van De Belt TH, Faber MJ, Knijnenburg J, Nelen WL, Kremer JA. Wikis as an opportunity to improve patient participation in developing information leaflets: a demonstration project in infertility patients. 2011 Presented at: Medicine 2.0; Sept 16-18, 2011; Stanford, CA, USA URL: http://www.medicine20congress.com/ocs/index.php/med/med2011/paper/view/582

Edited by G Eysenbach; submitted 06.10.11; peer-reviewed by M Laurent, S Gupta; comments to author 09.10.11; revised version received 10.11.11; accepted 18.11.11; published 08.12.11.

Please cite as:
Archambault PM
WikiBuild: A New Application to Support Patient and Health Care Professional Involvement in the Development of Patient Support Tools
J Med Internet Res 2011;13(4):e114
URL: http://www.jmir.org/2011/4/e114/
doi:10.2196/jmir.1961
PMID:22155746

©Patrick Michel Archambault. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Guest Editorial

eHealth Literacy 2.0: Problems and Opportunities With an Evolving Concept

Cameron Norman[1], PhD

CENSE Research + Design, Toronto, ON, Canada

Corresponding Author:
Cameron Norman, PhD
CENSE Research + Design
757-155 Dalhousie Street
Toronto, ON, M5B2P7
Canada
Phone: 1 4168543805
Email: cameron.norman@utoronto.ca

Abstract

As the use of eHealth grows and diversifies globally, the concept of eHealth literacy – a foundational skill set that underpins the use of information and communication technologies (ICT) for health – becomes more important than ever to understand and advance. EHealth literacy draws our collective attention to the knowledge and complex skill set that is often taken for granted when people interact with technology to address information, focusing our attention on learning and usability issues from the clinical through to population health level. Just as the field of eHealth is dynamic and evolving, so too is the context where eHealth literacy is applied and understood. The original Lily Model of eHealth literacy and scale used to assess it were developed at a time when the first generation of web tools gained prominence before the rise of social media. The rapid shifts in the informational landscape created by Web 2.0 tools and environments suggests it might be time to revisit the concept of eHealth Literacy and consider what a second release might look like.

(J Med Internet Res 2011;13(4):e125) doi:10.2196/jmir.2035

KEYWORDS

eHealth literacy, measurement, consumer eHealth, social media

Introduction

As the use of eHealth grows and diversifies globally, the concept of eHealth literacy – a foundational skill set that underpins the use of information and communication technologies (ICT) for health – becomes more important than ever to understand and advance [1]. eHealth literacy draws our collective attention to the knowledge and complex skill set that is often taken for granted when people interact with technology to address information, focusing our attention on learning and usability issues from the clinical through to population health level. Just as the field of eHealth is dynamic and evolving, so too is the context where eHealth literacy is applied and understood. The original Lily Model of eHealth literacy [1] and the scale used to assess it [2] were developed at a time when the first generation of web tools gained prominence before the rise of social media. The rapid shifts in the informational landscape created by Web 2.0 tools and environments suggests it might be time to revisit the concept of eHealth Literacy and consider what a second release might look like.

This issue of the *Journal of Medical Internet Research* (JMIR) provides examples of the diverse means in which the concept of eHealth Literacy has been applied, introducing challenges and presenting opportunities for understanding the evolution of the concept in the age of eHealth and mHealth. These challenges and opportunities will now be discussed in light of four papers published in this issue of JMIR [3-6].

The eHealth Literacy Concept and eHEALS

The eHealth literacy concept, model and related measurement scale, the eHealth Literacy Scale (eHEALS) [1, 2], originated from work that Harvey Skinner and I were doing on ICT-facilitated health promotion with youth and youth workers in the late 1990's and early 2000's [7 -11]. At the time, those under the age of 25 were among the most prolific and creative users of ICT's and thus, provided the ideal population to study the skill set required to access and fully engage with what became known as eHealth [12]. The concept of eHealth literacy was born of repeated observation in our research and health promotion practice that there was a noticeable gap between

XSL•FO
RenderX

consumers' absolute use of technology and the functional adoption of that technology into useful information finding and problem-solving. The eHEALS has been used in a variety of settings, with diverse population groups and has been translated into multiple languages [13-16]. The 8 or 10-item measure of eHealth Literacy continues to perform consistently across settings and populations.

eHealth Literacy in This JMIR Issue

The theory and measurement of eHealth literacy is not without challenges and the papers in this issue highlight some of them, while presenting opportunities for further learning.

The work by Stellefson and colleagues [3] looked at the state of eHealth literacy research with young people, drawing on the original study and model of eHealth literacy by Norman and Skinner [1,2] and also the definition of health literacy posed by the U.S. Institute of Medicine [17]. The paper demonstrates how things can get conflated when looking at the literacy issues within an eHealth context. The original Lily Model – referred to by the authors in the introduction – posits that eHealth literacy is a form of meta-literacy, combining many different literacy skills beyond just health literacy or numeracy. To focus solely on just one or two aspects of literacy within the model when assessing how it manifests in practice is problematic when making claims about eHealth literacy as a whole given that the concept is intended to represent a set of integrated skills. eHealth literacy operates as part of a learning system, whereby the component parts operate as a whole and not in a means that is easily amenable to subdivision.

The study by Chan and Kaufman [6] illustrate this complexity while expanding the scope of how eHealth literacy is assessed in the practice of eHealth use. The authors look at the concept of eHealth literacy as it is expressed in the practice of information seeking and contribution to interactive discussions by looking at the task demands required to fully engage with eHealth. They propose a framework for characterizing the task demands associated with eHealth use and in doing so extend the eHealth literacy model in light of practice, offering to fill the gap between the theory and the measurement of the concept. While the eHEALS was cited in the paper, none of the selected studies used the eHEALS. Furthermore, the study's inclusion criteria included papers that had "at least one aspect of eHealth literacy accounted for in the Norman and Skinner [1] definition used within this review". In taking this approach, there is a risk that eHealth literacy is reduced to a set of interchangeable skills without attention to how they combine. Indeed, the argument posed when the model was created was that eHealth literacy was the combined features of the six forms of literacy or petals in the Lily Model, not a subset of them.

Xie [5] took a different approach by looking at the eHEALS items along with measures of learning styles, preferences and general knowledge. The study focused on a population that has high needs for information, potentially greater isolation for informational resources, and a perceived lower familiarity with ICT's. By looking at the skill set of eHealth literacy within a larger learning context, Xie reminds us that learning – no matter what the subject matter – is highly contextual, often social, and dependent upon learning styles and opportunities to connect with others.

Context is also an issue with the eHEALS itself, as the work by van der Vaart and colleagues [4] (perhaps unintentionally) introduce in their critique of the validity of the eHEALS. At the outset, the eHEALS was designed to be easy to use and administer in response to the expressed needs of health professionals who said they would not use a long instrument in practice. After three years of development, an 8 and 10-item version of the eHEALS was created to address research and health practitioner needs. The eHEALS was put through a rigorous testing process to explore the internal consistency reliability and validity of the instrument. In the initial studies that contributed to the development of the eHEALS, both reliability and validity scores were high, indicating that the measure was suitable to use. Since its initial testing and the publication of the eHEALS in JMIR, the eHEALS has been translated into multiple languages and employed with a diverse population stream from Chinese children [11] through to older adults [18]. While the results have been consistently positive, there are issues with the way that the concept of eHealth literacy has been measured.

The research by van der Vaart and colleagues [4] questions the validity of the eHEALS in light of a weak correlation between eHealth literacy and Internet use. The findings by van der Vaart and colleagues could be due to cultural differences, measurement inconsistencies across studies, or it could be due to something else related to the evolving nature of eHealth altogether. When the eHEALS was first developed, the correlation between the two was high, so what has changed and why might this latest research reveal something different? One of the principal reasons may be that the Internet has changed. When the eHEALS was first developed, the technology sector was in recovery from the dot-com bust and still seeking to develop itself. Social media hadn't been realized, nor had widespread mobile Internet use taken off. Today, social media and the mobile web are among the most popular uses of the Internet among consumers [19]. While the eHEALS remains a strong tool for assessing competency with Web 1.0 related technologies, its fit with social media is unclear and the eHEALS feels incomplete. It is possible that these qualities are coming out with regards to current Internet use patterns, which are different and build on the foundational skills that the eHEALS measures.

Such questions may be less about validity in a specific sense and more about the validity in a more generalized sense of Internet usage. Indeed, the positive coherence of the eHEALS in other studies and its psychometric robustness might suggest adding a social media interactive subscale rather than a change to the existing items, which would have significant consequences for the psychometric integrity of the instrument. Items could be developed that consider skills and tasks like:

- confidence in expressing oneself clearly in social interactions online
- ability to synthesize professional and non-professional advice
- comfort and skill in navigating through information obtained through a mobile device

XSL•FO
RenderX

- ability to use apomediaries to filter relevant and trustworthy information [20].

Conclusion

The fundamental collection of skills that comprise eHealth Literacy have not likely changed, but the contexts in which they are expressed in the dynamic realm introduced by social media have. This presents an opportunity for research and practice to consider the ways in which eHealth literacy can be understood and fostered. The papers presented in this issue of JMIR provide important additions to the growing field of study in eHealth literacy and offer a glimpse as to where the concept may evolve to as more evidence unfolds.

Conflicts of Interest

Dr. Norman created the initial eHealth Literacy concept and measurement tool, the eHealth Literacy Scale (with Harvey Skinner).

References

1. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
2. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
3. Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA. eHealth Literacy Among College Students: A Systematic Review With Implications for eHealth Education. J Med Internet Res 2011;13(4):e102 [FREE Full text] [doi: 10.2196/jmir.1703] [Medline: 22155629]
4. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
5. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]
6. Chan CV, Kaufman DR. A Framework for Characterizing eHealth Literacy Demands and Barriers. J Med Internet Res 2011;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
7. Norman CD, Maley O, Li X, Skinner HA. Using the internet to assist smoking prevention and cessation in schools: a randomized, controlled trial. Health Psychol 2008 Nov;27(6):799-810. [doi: 10.1037/a0013105] [Medline: 19025276]
8. Norman CD, Skinner HA. Engaging youth in e-health promotion: lessons learned from a decade of TeenNet research. Adolesc Med State Art Rev 2007 Aug;18(2):357-69, xii. [Medline: 18605651]
9. Skinner H, Biscope S, Poland B, Goldberg E. How adolescents use technology for health information: implications for health professionals from focus group studies. J Med Internet Res 2003 Dec 18;5(4):e32 [FREE Full text] [doi: 10.2196/jmir.5.4.e32] [Medline: 14713660]
10. Skinner HA, Maley O, Norman CD. Developing internet-based eHealth promotion programs: the Spiral Technology Action Research (STAR) model. Health Promot Pract 2006 Oct;7(4):406-417. [doi: 10.1177/1524839905278889] [Medline: 16840770]
11. Skinner HA, Maley O, Smith L. Morrison M: New frontiers: Using the Internet to engage teens in substance abuse prevention and treatment. In: Monte P, Colby S, editors. Adolescence, alcohol, and substance abuse: Reaching teens through brief interventions. New York: Guilford; 2001.
12. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
13. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
14. Chan CV, Matthews LA, Kaufman DR. A taxonomy characterizing complexity of consumer eHealth Literacy. AMIA Annu Symp Proc 2009;2009:86-90. [Medline: 20351828]
15. Cheng HM, Norman CD, Koo M. Psychometric evaluation of Chinese version of the eHealth Literacy Scale (eHEALS) in school age children. International Electronic Journal of Health Education forthcoming.
16. Miller VM. Poor eHealth literacy and consumer-directed health plans: a recipe for market failure. Am J Bioeth 2007 Nov;7(11):20-2; discussion W1. [doi: 10.1080/15265160701638587] [Medline: 18027293]
17. Institute of Medicine. Health literacy: A prescription to end confusion. Washington, DC: Institute of Medicine; 2004.
18. Xie B. Experimenting on the impact of learning methods and information presentation channels on older adults' e-health literacy. Journal of the American Society for Information Science and Technology 2011;62(9):1797-1807. [doi: 10.1002/asi.21575]
19. Neilson. State of the Media: The social media report Q3 2011. New York, NY: Neilson; 2011.
20. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]

XSL•FO
RenderX

Edited by G Eysenbach; submitted 21.12.11; this is a non–peer-reviewed article;accepted 22.12.11; published 23.12.11.

Please cite as:
Norman C
eHealth Literacy 2.0: Problems and Opportunities With an Evolving Concept
J Med Internet Res 2011;13(4):e125
URL: http://www.jmir.org/2011/4/e125/
doi:10.2196/jmir.2035
PMID:22193243

©Cameron Norman. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies

Julia EWC van Gemert-Pijnen[1*], PhD; Nicol Nijland[1*], PhD; Maarten van Limburg[1], MSc, BEng; Hans C Ossebaard[2], MA; Saskia M Kelders[1], MSc; Gunther Eysenbach[3], MD, MPH, FACMI; Erwin R Seydel[1], PhD

[1]Department of Psychology, Health and Technology/Center for eHealth Research and Disease Management, Faculty of Behavioural Sciences, University of Twente, Enschede, Netherlands

[2]National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands

[3]Centre for Global eHealth Innovation, Department of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Julia EWC van Gemert-Pijnen, PhD
Department of Psychology, Health and Technology/Center for eHealth Research and Disease Management
Faculty of Behavioural Sciences
University of Twente
Drienerlolaan 5
PO Box 217
Enschede, 7500 AE
Netherlands
Phone: 31 534896050
Fax: 31 534892388
Email: j.e.w.c.vangemert-pijnen@utwente.nl

Abstract

Background: Many eHealth technologies are not successful in realizing sustainable innovations in health care practices. One of the reasons for this is that the current development of eHealth technology often disregards the interdependencies between technology, human characteristics, and the socioeconomic environment, resulting in technology that has a low impact in health care practices. To overcome the hurdles with eHealth design and implementation, a new, holistic approach to the development of eHealth technologies is needed, one that takes into account the complexity of health care and the rituals and habits of patients and other stakeholders.

Objective: The aim of this viewpoint paper is to improve the uptake and impact of eHealth technologies by advocating a holistic approach toward their development and eventual integration in the health sector.

Methods: To identify the potential and limitations of current eHealth frameworks (1999–2009), we carried out a literature search in the following electronic databases: PubMed, ScienceDirect, Web of Knowledge, PiCarta, and Google Scholar. Of the 60 papers that were identified, 44 were selected for full review. We excluded those papers that did not describe hands-on guidelines or quality criteria for the design, implementation, and evaluation of eHealth technologies (28 papers). From the results retrieved, we identified 16 eHealth frameworks that matched the inclusion criteria. The outcomes were used to posit strategies and principles for a holistic approach toward the development of eHealth technologies; these principles underpin our holistic eHealth framework.

Results: A total of 16 frameworks qualified for a final analysis, based on their theoretical backgrounds and visions on eHealth, and the strategies and conditions for the research and development of eHealth technologies. Despite their potential, the relationship between the visions on eHealth, proposed strategies, and research methods is obscure, perhaps due to a rather conceptual approach that focuses on the rationale behind the frameworks rather than on practical guidelines. In addition, the Web 2.0 technologies that call for a more stakeholder-driven approach are beyond the scope of current frameworks. To overcome these limitations, we composed a holistic framework based on a participatory development approach, persuasive design techniques, and business modeling.

Conclusions: To demonstrate the impact of eHealth technologies more effectively, a fresh way of thinking is required about how technology can be used to innovate health care. It also requires new concepts and instruments to develop and implement technologies in practice. The proposed framework serves as an evidence-based roadmap.

XSL•FO
RenderX

(J Med Internet Res 2011;13(4):e111) doi:10.2196/jmir.1672

KEYWORDS

eHealth; design; participation; implementation; evaluation; multidisciplinary approach; Health 2.0; Wiki; e-collaboration

Introduction

The impact of eHealth technologies is sometimes questioned because of a mismatch between the postulated benefits and actual outcomes. A lack of evidence about the distinct effects of eHealth technologies on health and health care is apparent [1-4]. Health care professionals are often skeptical and show little support for eHealth because technology does not seem to work for them or the benefit of their patients [5]. As a result, eHealth technologies often face adoption problems.

What could explain this mismatch? We know from research and the literature [1,2,4] that inadequate reimbursement and legislation can slow down the pace of innovation. Investors need to have trust before they can finance eHealth projects [2]. Apart from economic trust, a complex innovation needs coordination and communication [6], especially in the case of chronic disease management, where a variety of stakeholders are involved. Introducing eHealth technologies into the health care system requires careful coordination and communication among health care professionals, patients, informal caregivers, end users, and others. This is exactly what seems so hard to realize in practice. The same goes for project management; the precise definition of scope and objectives of the eHealth technology, the casting of participants, and the timely allocation of well-defined powers (eg, recourses and opinion leaders) and responsibilities are often not well defined beforehand. In day-to-day health care practice, these components are often present only on a superficial level, or not at all. In this situation, a lack of coordination and management deeply affects the outcomes from eHealth technologies research. Conversely, post hoc analysis does not, or cannot, account for the clouding of possible effects due to these important factors.

Another cause for the supposed low impact of eHealth technologies is the peripheral position of the users. eHealth technologies are often developed with only a marginal level of engagement from the (end) user. This lack of human centeredness explains the incidence of usability problems [7-9], or high attrition rates [10-18]. People simply stop using technologies that do not correspond in any way with their daily lives, habits, or rituals. In the end, the use of new technologies appears to be time consuming and frustrating for all those involved. In this way, technology-driven approaches result in “high tech-with-a-low impact” eHealth technologies [19-22].

All these confounding factors are not inextricably tied up with technology. Rather, avoiding them would reveal the real impact of eHealth technologies. The way in which technology is being designed to improve health care needs rethinking. The approaches that are being used to develop eHealth technologies are not productive enough to create technologies that are meaningful, manageable, and sustainable.

The development of eHealth technologies involves more than simply designing a product or service, and includes more than merely procuring stand-alone medical devices. We recognize the social dynamic and significance of eHealth technologies and their potential for improving health care. Creating a new technology often forces us to clarify how the process of health care delivery actually runs—for example, who the key stakeholders are and how payment is organized. It also illustrates the interdependencies between technology, people, their sociocultural environment, and the infrastructural organization of health care. Ideally, all stakeholders should be aware of these complex relationships [23].

In the wake of health 2.0 and medicine 2.0 initiatives [24, 25], a growing number of studies have emphasized the importance of a participatory development process involving (end) users—and other stakeholders such as payers, decision makers, insurers, and government officials—to increase the uptake of eHealth technologies [24]. Yet, in current perspectives on and definitions of health 2.0 [25], the role of stakeholders is not often addressed, nor is the potential of eHealth technologies to create infrastructures for better, cheaper, and easier-to-get health care services.

As long as the need to create a better fit between technological, human, and contextual factors continues to go unaddressed, the uptake and impact of eHealth technologies will remain at the very least poor, and at best undecided [4,26-28]. Therefore, we believe that a *holistic* approach is needed. Holistic means that we emphasize the importance of the whole and the interdependence of its parts, and avoid separate analysis of its parts. Such an approach would account for the issues of finance, management and technology when designing, implementing, and evaluating eHealth technologies. It constructs a productive fit through the integration of persuasive and human-centered design principles and business modeling. The urgent need for a holistic perspective to overcome the obstacles that stand in the way of the uptake of eHealth technologies has already been recognized [6,28].

The aim of this viewpoint paper is to boost the uptake and impact of eHealth technologies by advocating a holistic development approach. To this end, we undertook a critical appraisal of existing eHealth frameworks. First, we tried to identify the constituent elements of the framework: the target groups, the goals related to the development, implementation, and evaluation of eHealth technologies, the theoretical backgrounds, the visions on eHealth, and strategies or principles to increase the uptake and impact of eHealth technologies. In particular, we evaluated the extent to which the frameworks aim to realize a fit between human, organizational, and technological factors. Second, based on the outcomes of the review and supported by current knowledge on eHealth technologies development, we present the working principles for the holistic development process of eHealth technologies. And third, we build these principles into a holistic framework for developers, researchers, and decision makers. This holistic framework intends to guide the development of eHealth

XSL•FO
RenderX

technologies. It already does so in three of our case studies in infection management, dermatology, and diabetes. The roadmap represents our current view on the development of eHealth technologies. It is a dynamic framework and we also publish it as a wiki for collaborative use (http://ehealthwiki.org).

Review of Existing eHealth Frameworks

Selection Procedure

We searched the literature through the electronic databases PubMed, ScienceDirect, Web of Knowledge, PiCarta, and Google Scholar. Journal indexes were searched. Examples of journals searched are the *Journal of Medical Internet Research*, *International Journal of Medical Informatics*, *Telemedicine and e-Health*, *Journal of Telemedicine and Telecare*, and *Journal of the American Medical Informatics Association*. Using a snowball and cross-referencing methodology, we included relevant cited and related articles.

We included eHealth frameworks based on the following inclusion criteria:

1. The paper must be published in a peer-reviewed journal.
2. The paper must either describe an eHealth theory, perspective, framework, or model, or contain a literature review. We particularly sought frameworks that provide a set of guiding principles for improving the development, uptake, and impact of eHealth technologies. A framework is considered as a set of (1) principles: assumptions, constructs, quality criteria, and ideas that guide research and development, and/or (2) strategies: hands-on guidelines, design heuristics, and methods to assist the development process, and/or constructs or criteria that have to be met to increase the quality of eHealth technologies (definition based on Kaufman et al [29]).
3. The proposed framework must propose quality criteria for the design, implementation, and evaluation of eHealth technologies and must account for multilevel factors of a human, technical, environmental, or organizational nature.
4. The title of the journal paper must include at least one of the following search terms: *eHealth* or similar terms, such as telemedicine, telecare, telehealth, health information systems/technology, or interactive health communication applications; AND *development* AND/OR *design*, AND/OR *implementation*, AND/OR *evaluation*, AND *framework*, AND/OR *quality*, AND/OR *success* (in terms of improved or innovated health care referring to cost benefits, health outcomes, behavioral outcomes, or care organization).

We identified 60 journal papers (see Multimedia Appendix 1) based on the search criteria used. Journal papers with a general focus that described merely the potential of eHealth (16 papers, general) were excluded from the analysis. We selected 44 for a full review. From these, we excluded those papers that did not describe a framework providing hands-on guidelines, or quality criteria for the design, implementation, and evaluation of eHealth technologies (28 papers, nonframeworks). We did not make any restrictions regarding the kinds of technologies used. From the results retrieved, 16 eHealth frameworks (see Table 1, Multimedia Appendix 1, Multimedia Appendix 2, and Multimedia Appendix 3) were identified that matched the inclusion criteria. Multimedia Appendix 1 shows the included and excluded eHealth journal papers.

Table 1. eHealth frameworks that matched the inclusion criteria.

Framework	Corresponding author
fr.1	Esser & Goossens [104]
fr.2	Catwell & Sheikh [23]
fr.3	Yusof et al. [28]
fr.4	Hamid & Sarmad [50]
fr.5	Pagliari [48]
fr.6	Kaufman et al. [29]
fr.7	Dansky et al. [6]
fr.8	Van der Meijden et al. [30]
fr.9	Shaw [27]
fr.10	Kazanjian & Green [49]
fr.11	Kushniruk [60]
fr.12	Hebert [33]
fr.13	Eysenbach [117]
fr.14	Eng et al. [51]
fr.15	Jai Ganesh [52]
fr.16	Kukafka et al. [26]

XSL•FO
RenderX

Results

The first objective of the present review was to identify the strategies that are proposed for addressing the uptake and impact of eHealth technologies. Second, we wished to know how far these strategies draw on a holistic approach that strives to accomplish a fit between the human, organizational, and technology aspects. Since 1999, several eHealth frameworks have been published that describe a vision on how to increase the impact of eHealth technologies. Multimedia Appendix 2 presents the target groups and goals of the frameworks, the theoretical foundation, and definitions of eHealth technology that underpin the frameworks. Multimedia Appendix 3 presents the strategies and principles that are considered essential for the development of eHealth technologies, as well as the proposed evaluation methods.

Target Group

To whom are the frameworks meant to be applied and who is involved in carrying out the tasks (development, research, and employment) that have to be accomplished? From Multimedia Appendix 2 (target groups), it is clear that the frameworks are aimed at different target groups. These target groups vary from *single* groups—designers (framework fr. 1, 2), decision makers (fr. 10), and health planners (fr. 16)—to *multiple* groups—researchers and others (fr. 7), researchers and practitioners (fr. 3), researchers and developers (fr. 5, 6), developers, health care providers, purchasers, consumers, and policy makers (fr. 14).

The frameworks are usually expert driven; that is, they are meant for experts such as designers or researchers. However, these target groups are rarely specified, so the type of design professional or researcher that belongs to them remains unclear. An exception to this rule are frameworks 5 and 10, where health service researchers (fr. 5) are targeted, as well as policy makers and administrative developers of information systems (fr. 10). Quite a few authors do not specify the target group that their frameworks (fr. 4, 8 ,9, 11–15) are supposed to serve.

Six of the frameworks (fr. 3, 5 , 6, 7, 10, 14) address multiple target groups, although it is not clear what kind of tasks the different groups have to carry out in the subsequent development process. These frameworks provide, for instance, guidelines for evaluation as part of the development process. But it is not clear who is responsible for what kind of task. No specification has been provided for those (by discipline) who are involved in *producing* the eHealth technology and those who are involved in the *deployment* of the eHealth technology.

Goals

What do the authors want to achieve with the proposed frameworks? As shown in Multimedia Appendix 2, all of the frameworks aim to improve either the *uptake* (eg, implementation and adoption of eHealth technologies) or the *impact* (eg, effectiveness of eHealth technologies), or both. The frameworks are supposed to assist the target groups in the development of eHealth technologies via checklists, guidelines, and criteria. Frameworks that aim to enhance the uptake of eHealth technologies (fr. 1, 3–5, 7, 8, 10, 13–16) can be used for *formative (process-driven) evaluation purposes* to assess user acceptance and satisfaction, widespread adoption, or implementation (eg, infrastructure and resources) of eHealth technologies. The frameworks that aim to enhance impact (fr. 1–3, 5–8, 12, 14) can be used for *summative (performance-driven) evaluation purposes* to assess the potential of eHealth technologies to enhance the quality of health care, benefits, performance, and effectiveness (eg, health outcomes and cost reductions). Two frameworks (fr. 8, 12) aim to enhance the *success* of eHealth technologies. The term success is used by the authors in different ways. Van der Meijden et al [30] (fr. 8) refer to the six dimensions of success defined by DeLone and McLean [31,32]: system, service and information *quality*, to user *acceptance*, and individual and organizational *impact*. Hebert [33] (fr. 12) refers to Donabedian's [34] quality-of-care measures: structure, process, and outcome. The frameworks that aim to increase both the uptake and impact (fr. 1, 3, 5, 7, 8, 14) have more potential to create a perfect fit between human, organizational, and technology factors. Remarkably, only two frameworks guide the decision-making process via scientific evidence (fr. 2, 10), while one channels the investment of eHealth technologies (fr. 12). Some frameworks (fr. 9, 11) do not explicitly state any goals.

Foundation

What theories or models underpin the frameworks? What empirical evidence is this grounded in? The frameworks are based on a combination of models, theories, and literature review studies (fr. 1–5, 8, 10–13, 15, 16), and some are validated by experts (fr. 1) or tested via empirical research (fr. 3, 9, 12). The authors of the frameworks (fr. 4, 16) argue that the development of eHealth technologies should be grounded in multidisciplinary theories such as behavioral and sociocognitive theories and those linked to innovation and diffusion. Two frameworks are based on regulations (the Health Insurance Portability and Accountability Act [6], fr. 7) and institutional regulations that make health communication programs work (National Cancer Institute [35], fr. 14).

The theoretical foundations of current frameworks stem from *human–technology interaction* models based on software engineering principles and behavioral theories; *health service* models for quality of health care; and *innovation diffusion* models (Multimedia Appendix 2: foundation). Given the complexity of health care, some authors argue that more contingency-driven models are needed to address the organizational and environmental aspects that influence the uptake of eHealth technologies: sociotechnical and contextual aspects (fr. 2), the IT-Organization Fit model (fr. 3), diffusion theories (fr. 4), health services evaluation methods (fr. 5, 10, 12), and cross-theoretical integration of behavior models (fr. 16).

The *human–technology interaction* models (fr. 1–6, 8, 11, 16) are aimed at achieving user centeredness for eHealth technologies, which is considered to be a prerequisite for the acceptance of eHealth technologies in practice. Examples are software design models [36-38], information system success models [31], and technology acceptance models [39-41]. These models focus on the factors that influence usability, acceptance,

XSL•FO
RenderX

or adoption of eHealth technologies. Framework 1 is a *human–media interaction* framework based on the media richness theory to support the patient–caregiver interaction (based on Miller [42]).

Frameworks 12 and 13 are based on *service-quality* models such as Donabedian's [34] quality-of-care measures. Health technology assessment (fr. 12) was used as an approach to assess the value of eHealth technologies in practice, and in the World Health Organization strategies for (re)designing health care systems (fr. 14).

The frameworks that highlight the contextual aspects that are important for the integration and operation of eHealth technologies in the health care context are founded in *innovation diffusion* models such as the Precede-Proceed Model (fr. 16), social-cognitive theory, diffusion of innovation theory (fr. 15, 16), IT-Organization Fit Model (fr. 3), and acts/legislation such as the Health Insurance Portability and Accountability Act (fr. 7, 10).

Through literature reviews we identified aspects that influence the success of eHealth technologies (fr. 8, 13) and aspects that are critical for evaluation during the development of eHealth technologies (fr. 2–6, 10, 12, 16). The results were used to ground the development approach (stages from ideation to rollout) or to formulate criteria for the evaluation of research activities related to the development of eHealth technologies. Empirical evidence for the frameworks stem from expert validation (fr. 1) to assess the relevance of the frameworks or from pilot testing in practice (fr. 3, 9, 12) to assess the utility of the framework.

eHealth Definition and Technology Focus

What definitions of eHealth were used as a basis for the frameworks? What kinds of technologies did the frameworks focus on? Some authors use their own definitions of eHealth (fr. 1, 5, 8, 15; see Multimedia Appendix 2). Framework 1 uses a definition for telemedicine and refers to the use of information and communication technologies for the exchange of medical information in a clinical setting, aimed at a specific technology, teleconsultation. Framework 5 refers to medical informatics and health services research as a synonym for eHealth, related to health technology assessment and health systems research. The definition of framework 8 refers to health information systems: general patient care information systems in hospital settings or specific care information settings (intensive care unit). Framework 15 defines eHealth as the use of electronic information and communication technology to promote health or improve health care. The authors (fr. 15) explicitly mention that the infrastructure of an eHealth program consists of three components: human, technical, and medical.

The other definitions come from researchers in the field of eHealth. These definitions are in a certain way related to the technologies the frameworks focus on. For example, the definition from the Institute of Medicine is used (fr. 10) to classify health information systems and the definitions of telehealth by Field [43] and Reid [44] are used to describe the use of technologies in rural areas or in cases where social or cultural barriers separate the participants (fr. 12). Four frameworks (fr. 2, 4, 7, 14) have a wider focus referring to multiple technologies and modalities for the organization and delivery of health care services and information. Framework 2 [23] used the definition of Eysenbach [45] (see Multimedia Appendix 2). The authors of framework 2 argue that "...the definition of eHealth should encompass the full spectrum of ICTs [information and communication technologies], whilst appreciating the context of use and value they can bring to society" [28] is aimed at better service utilization via eHealth technologies in general. It uses a description of eHealth from Canada's Health Informatics Association (defined in Oh et al [46]) to connect providers, patients, and governments; to educate and inform health care professionals, managers, and consumers; to stimulate innovation in care delivery and health system management; and to improve the health care system.

The authors of framework 7 expand on the evolution of eHealth, changing from a one-way system to wireless technologies and online communities using Web 2.0 technologies. The authors state that eHealth is revolutionizing health care, resulting in new models for eHealth development:

> *eHealth has moved from an acute-care orientation to prevention and disease management, from an individual focus to a population focus, and from an institutional setting to communities and cyberspace. Concomitantly, models of healthcare delivery have evolved from being physician and clinician driven, to patient-centered care models that are based on participative decision-making [6].*

Framework 14 uses a definition from Robinson et al [47] for interactive health communications, which is the focus of their framework (see Multimedia Appendix 2). They posit that they do not focus on eHealth technologies that focus exclusively on logistics or clinical data. In some cases, no definitions on eHealth were reported (fr. 3, 6, 9, 11, 13, 16).

Strategies and Principles for eHealth Research & Development

The frameworks propose different strategies and principles for the development of eHealth technologies (presented in Multimedia Appendix 3). Almost all of the frameworks plead for a *multidisciplinary development approach*, *continuous and systematic evaluation* during development, and *robust methods* for formative and summative evaluations to realize technologies that are aligned with the needs of their users and environmental aspects.

A Multidisciplinary Development Approach

Several frameworks (fr. 2–5, 7, 9, 10, 14–16) posit that a *multidisciplinary development approach* (see Multimedia Appendix 3) is important when developing eHealth technologies. A multidisciplinary approach is considered as the involvement of different disciplines in the development of eHealth technologies referring to a team of various experts that guide the development, or the involvement of various stakeholders that can be affected by the use of the eHealth technologies.

Some frameworks advocate a user-centered design approach that takes into account the needs of the end users (patients and/or

XSL•FO
RenderX

health care providers) during the development process (fr. 1, 4, 6, 8, 11–13). Others propose a comprehensive overall approach that addresses the importance of involving different stakeholders (patients, clinicians, managers, information technology providers, the health care organization, etc) in the development process (fr. 2, 3, 7, 10, 14–16) in order to document the complex relationships between political, social, organizational, and technical worlds (fr. 2); ensure that different contexts and visions are taken into account (fr. 3, 7, 9, 10); identify the values and concerns different stakeholder have (fr. 16); or develop sustainable eHealth technology programs (fr. 15).

Pagliari [48] (fr. 5) and Kazanjian and Green [49] (fr. 10) argue that a multidisciplinary development *team of experts* is needed to maximize the potential of eHealth. Such a team should consist of a "wider constituency of disciplinary experts including social, management, and legal scientists, all of whom have a stake in the field" (fr. 5) and "a number of disciplinary perspectives, incorporating theories of epidemiology, sociology, economics, and systems science; and applies critical theory to health care evaluation" (fr. 10). Dansky et al [6] (fr. 7) state that a multidisciplinary team is needed to organize the development process to identify the staff and skills needed to implement eHealth technologies, and that roles and responsibilities should be identified in order to organize the research (data collection) and communications (involving stakeholders).

The participation of *stakeholder*s is viewed as essential to reflect on the values, drivers, and goals of the eHealth technologies to be developed. For example, Catwell and Sheikh [23] (fr. 2) argue that design teams need to have a thorough understanding of the stakeholders' needs, concerns, values, and beliefs, and define what the eventual system will be expected to provide:

> *The rich picture of the real world needs to be developed into a conceptual model so that stakeholders can reflect critically on the drivers, vision, and goals of the project and agree whether such a program of change is appropriate and feasible...It is important that the initial elicitation stage goes beyond functional and technical requirements and considers for example, accessibility, acceptability, and affordability issues [23].*

Hamid and Sarmad [50] (fr. 4) state that multiple stakeholders should be involved in the evaluation process. They state that one stakeholder, or a group with a common perspective, is the most important to be addressed: the user. This is a different view from the other frameworks in so far as other frameworks argue that a multiperspective view is needed to ensure that eHealth technologies fit with their users and other contingent variables (culture, organizational needs, etc). Eng et al [51] (fr. 14) refer to the statement of the Science Panel on Interactive Communication and Health, namely that four stakeholder groups should participate in order to improve the quality of eHealth technologies: consumers (patients, families, and caregivers), health care professionals and purchasers, developers, and policy makers. They argue that these stakeholders should participate at an early stage of developing the applications "to increase the probability of a favorable impact on health and quality outcomes." Jai Ganesh [52] (fr. 15) posits that a multidisciplinary collaboration is needed to increase acceptance (by consumers/patients and health care professionals) and "to establish joint ventures in the field of eHealth by inviting local or foreign partners to participate and to take equity stakes in the delivery of eHealth services." In this approach, it is important "to identify appropriate partners to specify appropriate technology and to find financing." The key players that should collaborate are the patients, practitioners (health care professionals), and health care service providers (see Multimedia Appendix 3). The aim of the collaboration is to bring together information technology experts, health professionals, lawyers, industry representatives, and others to ensure sustainable eHealth technologies.

Kukafka et al [26] (fr. 16) promote active participation via a participatory design approach to ensure that planners have a "structure in place to engage system end-users effectively from the start." They state that end users, management staff, and administrators should all be engaged in diagnosing the problems. "This process enables planners to expand their knowledge of the organization by identifying the values and subjective concerns key stakeholders have with existing systems and procedures." The authors do not specify what kind of stakeholders should be involved to facilitate multidisciplinary collaboration.

The authors of the frameworks that argue for a multidisciplinary development approach have different views on who should participate in the development process and what is actually meant by a multidisciplinary approach. The frameworks describe the actors that should participate in the development process either in terms of *disciplines*—clinical, human, social, organizational (environment), administrative (logistics), technical (industry), and political—or in terms of *stakeholder groups*—technology developers and health service researchers, clinicians/health care providers, payers, purchasers, policy makers, lawyers, and consumers/end users (patients, families, and caregivers).

Continuous and Systematic Evaluation

Multimedia Appendix 3 shows that some frameworks (fr. 2, 5, 6, 8, 11, 14) explicitly promote a continuous and systematic evaluation throughout the development of eHealth technologies to ensure that eHealth technologies are truly user informed, fit for context, of high quality, and of demonstrated value (fr. 5). These *process-driven* frameworks describe a cyclic, iterative research evaluation and development approach. The process starts with identifying the needs and goals of the intended users or stakeholders, compiling a requirements analysis of the design, prototyping, and implementation. Each stage is accompanied by its own set of evaluation research activities. For example, framework 5 provides different evaluation phases for research (formative and summative) from concepts to rollout. According to the author, 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 life cycle, from initial conception through rollout. Framework 11 presents formative and summative evaluation phases as well, from project planning to implementation. The authors of framework 14, Eng et al,

XSL•FO
RenderX

state that evaluation should be woven throughout all stages of the development process: conceptualization (formative evaluation; needs driven), design (prototyping), implementation and dissemination of product development (process evaluation of operationalization), and outcome evaluation (refinement). Evaluation is seen as crucial for the development of eHealth technologies, and the research activities for formative evaluations are related to each stage of the development process. Formative evaluation is considered to be a central iterative research activity that should be initiated during the early stages of development in order to assess the problems and needs of the various stakeholders. Van der Meijden et al [30] (fr. 8) argue that evaluation is often aimed at measuring the effects (summative), while neglecting the value of formative or process evaluation to improve a technology during development and implementation. Evaluation, in their view, should start *before* the technical development and has no fixed end because the technology fluctuates over time. The evaluation should include "multiple, selected periods of data collection and all stakeholders' points of view."

Frameworks 1, 3, 4, 7–10, and 12–16 can be considered as *quality-assessment frameworks*. These frameworks provide quality-assessment criteria, or evaluation dimensions and factors for the design, implementation, or success (ie, impact) of eHealth technologies. The authors of frameworks 3, 7, 9, and 16 state that it is important to address all aspects of health care that can be influenced by the use of eHealth technologies to ensure the fit between the technology, its users, and all of the organizational aspects. The evaluation dimensions influence each other in a temporal and causal way; this means that *fit* can be viewed in terms of strategic planning and strategic alignment (managing technology with organizational needs), as well as in terms of a fit between human and organizational needs and alignment between human needs and technology.

In fact, all of the quality-assessment frameworks address in a certain way contextual or environmental factors. Framework 1 presents criteria for teleconsultation specified to its users, the patient–provider communication, and the compatibility with the organizational context. Framework 4 is aimed at user centeredness, providing criteria for the fit between the system and its users, and the system's fit with the health care system. Framework 16 proposes critical assessment phases to determine organizational needs and goals amenable to technology (information technology solutions) and factors that influence human behavior. According to Yusof et al [28] (fr. 3), an evaluation of human, organizational, and technology aspects is required throughout the whole development cycle of planning, analysis, design, implementation, and operation and maintenance.

Robust Methods for Evaluation

The authors of several frameworks (fr. 2–6, 8–11, 14) argue that more rigorous evaluation approaches or methods are needed to assess the added value of eHealth technologies (see Multimedia Appendix 3). A mixed-methods approach is often proposed, combining qualitative methods (such as observations and interviews) with quantitative methods (workflow sampling or questionnaires) (fr. 2, 3, 5, 6, 8, 9, 11, 14). Van der Meijden et al [30] (fr. 8) argue that the integration of qualitative and quantitative data collection methods provides an opportunity to improve the quality of the results through triangulation, as the data from different sources complement each other to provide a more complete picture. "The integration of qualitative (observations, interviews) and quantitative (questionnaires, work sampling) data collection methods provides an opportunity to improve the quality of the results through triangulation" [30].

Formative process evaluation measures (fr. 2, 6) and longitudinal process studies (fr. 5) are recommended to demonstrate the acceptability and utility of new eHealth technologies, as well as the conditions for implementation that may influence their adoption. Summative methods are mentioned to evaluate the validity and efficacy of a system, such as randomized controlled trials (RCTs) (fr. 5, 6), or the overall impact (fr. 9, 11). The authors of frameworks 5 and 9 state that methods should go beyond the use of RCTs to evaluate the impact of eHealth technologies, because RCTs are seen as less well suited to evaluate the impact of eHealth interventions in a complex environment or to study the effect they have on the care delivery process. "An eHealth technology is not a drug and should not be evaluated as such," in Shaw's view [27] (fr. 9).

Catwell and Sheikh [23] (fr. 2) are of the opinion that "formative iterative evaluations using simple prototypes of the eHealth technology may be useful during the early stage of the development process to assist with communicating ideas, building a common understanding, agreeing to objectives, and securing stakeholder buy-in." Except for usability and prototyping instruments (fr. 2, 11), think-aloud methods for assessing usability (fr. 5), and the multiple methods presented by Yusof et al [28] (fr. 3) and Eng et al [51] (fr. 14), none of the frameworks present practical evaluation tools that are appropriate for the participation of various stakeholders in the development of eHealth technologies.

Potentials and Limitations of the Reviewed Frameworks

Potentials

The reviewed frameworks provide useful insights for the development of eHealth technologies that are user centered and fit for context: they provide process guidelines and indicators for creating eHealth technologies that are acceptable, affordable, and manageable. The added value of the frameworks lies in the multifactor approach: combining factors that support the adoption and implementation that are often underestimated in research (eg, fr. 8, 9, 16).

The integration of different models or theories is another added value, to ensure that eHealth technologies are feasible and sustainable. Furthermore, the comprehensive and integrative approach of some of the frameworks (eg, fr. 3, 7, 9, 15) are of interest to widen the contribution of eHealth technologies to innovate health care and to reduce societal problems (eg, aging or costs).

Most frameworks advocate a multidisciplinary development approach, involving collaboration among different stakeholders (eg, payers, technology providers, patients, and health care

XSL•FO
RenderX

professionals) and multiple methods (quantitative and qualitative) for assessing the process of technology development (formative evaluation) and the effects of eHealth technologies on health care (summative evaluations). One of the challenges expressed by the authors is the move from evaluations focusing exclusively on measuring outcome variables (via RCTs) to evaluations involving in-depth process data about the usage of the eHealth technologies in different care settings (hospitals or home-based care).

Limitations

Target Group

From the review we can conclude that the current frameworks have certain limitations. It is often not clear whether the frameworks are conceptual thoughts to provide insights and knowledge about the development of eHealth technologies or to use as a debating tool among researchers and others; or whether the frameworks should function as a practical guideline to assist developers. If target groups are reported, then it is not clear what kind of roles, tasks, or responsibilities these groups have during the development process. Another point is that the target groups are not specified. What kinds of disciplines or professions are representative of the researchers, designers, and other target groups? What level of cooperation is supposed between the different disciplines such as researchers, designers, and technical developers? And who is involved in the various stages from ideation to maintenance?

Goals

The frameworks aim to bring about the widespread diffusion and adoption of eHealth technologies, the implementation of eHealth technologies, or the improvement of the performance and effectiveness of eHealth technologies. It is often not clear what is meant by success, effectiveness, or performance. To realize the goals, different strategies and principles are presented. One of the concerns refers to the benefits or drawbacks of the frameworks, given the aims they strive for. It is unclear what the frameworks contribute toward increasing the uptake of eHealth technologies or innovating health care considering the wider social, political, or economic impacts of improvements in goal attainment.

Foundation

Evidence for the frameworks is based on scientific research or on literature reviews. A few empirical studies have been reported that help to ground the framework or demonstrate the benefits or drawbacks of the framework. It is often unclear how the theories or models that underpin the frameworks match the strategies reported. In general, the relationship between the visions on eHealth, proposed strategies, and research methods (formative and summative) are obscure, perhaps due to a rather conceptual approach that focuses on the rationale behind the frameworks rather than on practical guidelines.

eHealth Definition and Technology Focus

The frameworks address, in most cases, the evaluation of technologies that have already been defined (such as teleconsultation or health information systems), except for a few that do not focus on any specific form of technology in particular. The discussion about how to track down information about what kind of technologies (content or format) fit best with stakeholders' needs and values and care settings is not yet underway. Another limitation is the bias toward information systems rather than social or safety technologies. The use of social media for cocreation is beyond the scope of the frameworks.

Strategies and Principles for eHealth Research & Development

Although the importance of a multidisciplinary approach during the development of eHealth technologies is acknowledged by most frameworks, only a few authors have actually worked on incorporating this into the strategies for research and development within their frameworks. The frameworks propose stakeholder involvement during the development process, as part of the evaluation cycles. Most of the frameworks do not provide any insights into the identity of the stakeholders, sometimes referring to the end users, sometimes to developers and others (eg, health care professionals, providers, and government) that have a stake in the development of eHealth technologies. Referring to Yusof [28], it is not clear who participates (which stakeholders' perspective is going to be evaluated), when participation is needed, and by whom (at which phase in the system development life cycle), what the focus of the participation is (aspects or focus of the evaluation), and how participation can take place (methods of evaluation). In addition, the division of tasks between developers and stakeholders is not concise. In our view, the involvement of stakeholders is not restricted to the evaluation but extends to the full development process. Their involvement is important from ideation to the validation of business models.

Although the lack of contingency variables and the dominant focus on summative evaluations is reported as one of the main shortcomings of earlier models or frameworks, the current frameworks address the importance of contextual factors, but they do not incorporate these factors systematically into the frameworks' strategies. In fact, implementation is not considered as being interwoven within the development of eHealth technologies. Notably, no models are mentioned for collaborative development (ie, participatory development, cocreation, or value creation), incremental development, and sustainable implementation (business modeling). The critical point is that no clear information is given about the periods or the roles and responsibilities of different stakeholders in several stages of the development process, or about the focus of evaluation regarding the involvement of various stakeholders, or the methods for participation.

The authors of the frameworks argue for more rigorous evaluation methods that go beyond the use of experimental designs (RCTs) to evaluate the impact of eHealth technologies, because RCTs are less well suited to evaluate the impact of eHealth interventions in a complex environment. Rigorous qualitative studies combined with quantitative and process evaluation measures are recommended; the instruments that accompany this, however, are rarely reported. The participation of users and stakeholders is considered important, but the methods and instruments needed to guide this participatory

XSL•FO
RenderX

process are missing. The frameworks prescribe what should be done, but do not point to the instruments or tools to realize it. In fact, the greatest limitation of the frameworks is the lack of clear handles to support the development process; although the authors posit in their description of the frameworks the essential criteria for that, they did not translate this into their frameworks. For example, collaboration between the developers and the researchers is recommended, but no guidelines or prescribed activities are available for managing this type of development collaboration.

A Holistic Framework for the Development of eHealth Technologies

We believe that a comprehensive view on supporting health care by technology is needed to ensure that eHealth technologies are used effectively and efficiently. That is to say, that they realize their objectives and do so with optimal use of resources (time, money, and staff). We propose a holistic approach for the development of eHealth technologies. It is based on the outcomes of the reviewed frameworks, on empirical research, and on progressive insights obtained from discussing the framework with researchers (eHealth conferences).

Holism maintains that properties of individual elements in a complex system are taken to be determined by the relations they bear to other elements [53,54]. When applied to social theory this means that "each term owns its meaning to its relations with the others, so that they are all more or less closely inter-defined, and a change in the meaning of one term will have repercussions for all the rest" [55]. A holistic perspective on eHealth technologies has been advocated elsewhere, for instance by Dansky et al [6], Yusof et al [28], or Kukafka et al [26].

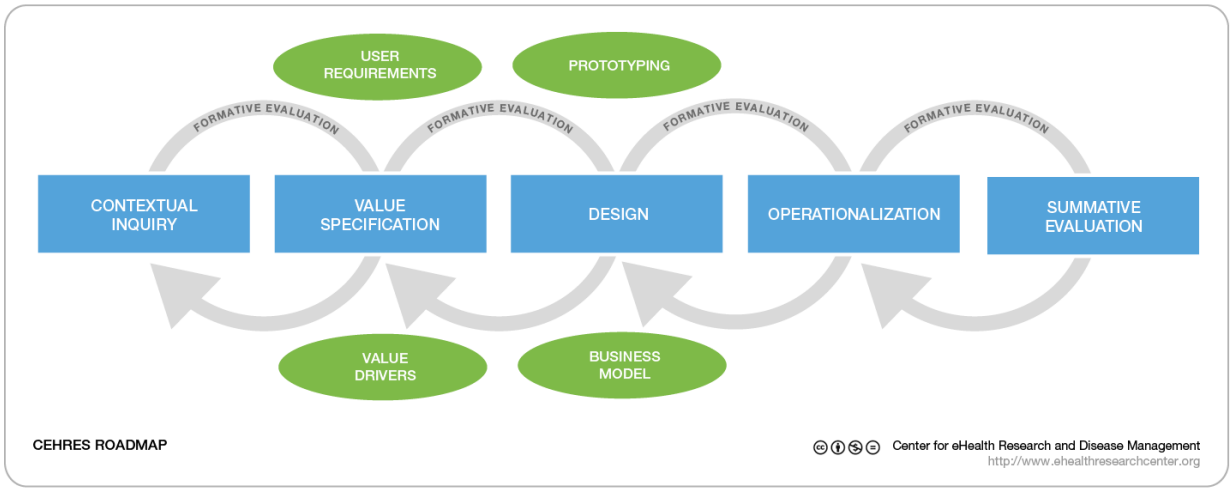
> *Without addressing the full range of factors, strategies to change behavior run the risk of being ineffective because they fail to recognize interdependencies between individual and organizational factors [26]*

For us it means that human characteristics, socioeconomic and cultural environments, and technology are considered to be closely connected to each other. eHealth technologies affect people in their daily lives. People always bring in their psychological makeup, their rituals and habits, and their social skills, which affect their personal and professional environment. This evidently affects their ability to interact with technology.

Strategies and Principles for a Holistic Development Approach

We introduce six working principles derived from the review of current frameworks, as well as from empirical research [7-9,14,21,56]. These principles are the groundwork for a holistic framework for the development of eHealth technologies. The framework and its related concepts are presented in Figure 1.

Figure 1. CeHRes Roadmap for the development of eHealth technologies.



1. eHealth Technology Development is a Participatory Process

EHealth technology development is a matter of cocreation; stakeholder participation is essential [25]. Stakeholders' involvement spans the full development process, starting from contextual inquiry and ending with summative evaluation (see Figure 1). Stakeholders can be considered as actors that have different roles in the development of eHealth technologies, from ideation to operationalization. Through their roles in identifying needs, or specifying critical issues for design and implementation, they help to create the technology [57,58]. Adequate project management needs to arrange for the participation of stakeholders and to identify their roles, tasks, and responsibilities.

2. eHealth Technology Development Involves Continuous Evaluation Cycles

Development is an iterative, flexible, and dynamic process resulting in concepts of the technology (from ideation to prototypes). These concepts need to be evaluated continuously. Evaluation as such is a cyclic, longitudinal research activity interwoven with all stages in the development process and as such without a fixed end (formative and summative evaluation; see Figure 1). By its usage, technology is reshaped. In the case of interactive technology, users reflect on the content or system

XSL•FO
RenderX

via feedback [59]. So, while technology evolves into action, evaluation follows on as a continuous, reflexive process that leads to matching human, organizational, and technology factors. This is noted in several of the reviewed frameworks [23,29,48,51,60]. Such evaluation takes place via formative cycles (see Figure 1) to reflect on the development process. Reflection is important to check for tacit understanding [61]. Stakeholders (including users) provide feedback and forward comments during the process [59]. They monitor their usage from their perspective.

3. eHealth Technology Development is Intertwined With Implementation

Implementation is often seen as a postdesign activity. In our view, the conditions for implementation must be taken into account right from the start (contextual inquiry and value specification; see Figure 1) [62]. Potential implementation issues such as limited resources (eg, time, staff, and money) or personal drawbacks (eg, skills, motivation, and anxieties) should be identified [63-66]. These issues should also be accounted for in the subsequent stages (design and operationalization; see Figure 1). In this way, the well-known pitfalls of stakeholder disregard can be avoided.

4. eHealth Technology Development Changes the Organization of Health Care

The development of eHealth technology in itself can be considered as the creation of new processes and infrastructures for health care delivery. It may reshape health care since it intervenes with traditional care characteristics such as the division of labor, or time- and place-dependant delivery [67]. This can be observed in today's shift from hospital-based care to home-based care, which requires new reimbursement and planning systems. Though underestimated in current research, this catalyst effect is inherent to eHealth technology development.

5. eHealth Technology Development Should Involve Persuasive Design Techniques

More and more patients wish to use technology for communicating and sharing personal information. They expect self-care technology to show understanding, to persuade them to do the right thing, or to provide rewards and appraisal for appropriate behavior [7,22]. However, the inherent capacities of technology as a persuasive medium for behavior change, information sharing, or self-management are often overlooked. Also, eHealth technologies often do not fit with the intended users' needs [7-9,14,68-70]. Particularly in the context of long-term care, it is important to develop technologies that can create bonding relationships with the end users. Motivation and social support are functionalities of eHealth technologies that belong to the most important persuasive drivers [20]. Via persuasive techniques, eHealth technologies can be designed to match user profiles, and to motivate or inspire patients to engage in self-management [71-75] (design; Figure 1).

6. eHealth Technology Development Needs Advanced Methods to Assess Impact

Several authors [1,2,4,6,62,76-87] note that a broader view is needed to assess the overall impact on health care. Both benefits and drawbacks have to be taken into account in terms of risks, ethics, performance, finance, or adherence. Impact is understood to evolve from an interaction between the technology, the person, and the context of usage. Interactions may be planned to be effective, to realize the outcome of the process. The process itself may be intended to be efficient, to use relatively few resources to achieve an objective. However, the constituents of this interaction are interdependent and mutually affect each other in a fuzzy manner: they may miss the target or take the hardest road. A holistic approach brings these elements together and targets the dynamics toward a desired, temporary situation that in the end is greater than its parts. Impact fluctuates over time and situations [30,88-90]; therefore, we need to have more advanced methods to assess the added value of eHealth technologies for health care and society. We need to understand *what* differences eHealth technologies can make in health care, *why* eHealth technologies make these differences, and *why* eHealth technologies may not have the expected impact [85,86,91]. Research on the impact of eHealth technologies is often done in clinical studies, mainly RCTs. This "gold standard" is often not suitable for identifying what works in practice. There are simply too many confounding factors that cannot be controlled or anticipated [83,92,93]. Moreover, impact in terms of organizational innovation cannot be measured in an RCT. For an RCT a certain degree of reduction is needed. This is exactly what influences the sustainability or effectiveness to be measured. Given these downsides, generalizing from their outcomes is at the least problematic. Some authors [30,48,60,80,84,85,94] have expressed the need for mixed methods using both quantitative and qualitative designs in order to better measure the uptake and impact of eHealth technologies.

The challenge lies in the integration of data collection from multiple sources, using a mixed-methods research design [95]. This implies a periodic evaluation during development rather than a before-and-after design, and advanced methods focusing on examining process variables (usage/dropouts of eHealth technologies) with methods measuring outcome variables (costs, health condition, or adherence to technology and interventions). The methods and instruments from the holistic framework will be described in a subsequent paper.

Holistic Framework: CeHRes roadmap

Framework, Target Group, and Goal

To visualize and outline our holistic approach we have created a framework and have presented a CeHRes roadmap (Figure 1). This roadmap serves as a practical guideline to help plan, coordinate, and execute the participatory development process of eHealth technologies. The framework is meant for developers (eg, technicians, designers, and health care professionals), researchers, and policy makers and for educational purposes (eg, students and health care providers). It also serves as an analytical instrument for decision making about the use of eHealth technologies.

XSL•FO
RenderX

eHealth Definition and Technology Focus

For the purpose of this paper, the term eHealth, or electronic health, refers to all kinds of information and communication technology used for supporting health care and promoting a sense of well-being. Within eHealth, a broad spectrum of technologies is used. These technologies include *Internet technologies*, such as informational websites, *interactive health communication applications* (ie, e-consultation, online communities, online health decision-support programs, and tailored online health education programs), *online health care portals*, and *electronic health records*. It also includes *mobile health communication programs*, and other advanced technologies such as *virtual reality programs* (ie serious gaming to stimulate exercise or 3-dimensional applications for the treatment of anxiety disorders), *home automation (domotics)*, sensor technology for independent living and remote monitoring, and *robotics*, the deployment of robots for assisting people with domestic tasks or to perform surgery.

The framework does not focus on the development or redesign of a specific technology; it should be used for all kinds of eHealth technologies with a scope broader than merely monitoring for medical purposes. In fact, our ultimate goal is to realize 5M-eHealth technologies that can support measurement (e-diagnose), monitoring (observation), mentoring (nudging), motivation (support), and management of data (automated integration of different data).

Foundation

The foundation of the framework is based on the aforementioned principles, reviews in the field of eHealth [2,3,22,96-99], progressive insights from eHealth research (see Multimedia Appendix 4), and multidisciplinary theories from psychology, communication, and human–computer interaction design.

In our view, integrating *persuasive technology design*, *human-centered design*, and *business modeling* provides the theoretical background for the development, evaluation, and implementation of eHealth technologies. As indicated by the authors of the frameworks that we reviewed, the development approach of eHealth technologies should be multidisciplinary in nature. Persuasive technology is the covering concept, referring to the use of technology to change people's attitudes and behavior [71-75]. The conception that technologies, especially interactive technologies, can persuade people to do the right thing at the right moment is rather new in the health care domain. Technology, for example, can simplify or guide people through a process of self-care management or provide social support through tailored feedback. In the domain of eHealth, we think that research into persuasive design techniques is needed to understand how technology can motivate or inspire healthy behaviors, how the technology fits with the needs of users (human-centered design) [68,100-104], and how technology can create new structures for health care delivery.

In addition, the participation of stakeholders, such as caregivers, insurers, or decision makers, influences the development of eHealth technologies. Their needs, concerns, values, and beliefs determine what the eventual technology should provide in order to realize the goals. To understand and guide the value-creation process in order to develop eHealth technologies that are affordable and interoperable with the health care system, innovation models are needed. Jai Ganesh [52] states that eHealth programs should be based on sound economic frameworks to deliver value for the investment of eHealth technologies. eHealth technologies require substantial financial investment [105]. The business case for eHealth technologies depends on the expectation of a return on investment. Nonetheless, we should focus not only on value in terms of money, since eHealth technologies may have value for life. *Business modeling* originated in commercial strategic management [61,106-109], focusing on the collaborative efforts of value creation with stakeholders. Stakeholders, the ones who affect or are affected by eHealth technologies [57,58], reflect on each other's values and weigh the importance of the values in terms of economic, behavioral, and psychological interests. This results in business models for the implementation of eHealth technologies. Concepts and techniques from business modeling help to identify critical factors for the implementation.

Research and Development Activities

Figure 1 depicts the development process and accompanied research activities. The research and development activities will be explained below.

Multidisciplinary Project Management

Ideally, the development process of an eHealth technology should start with multidisciplinary project management. Multidisciplinary project management facilitates the cooperation between those who are responsible for *producing* the technology (eg, technicians, designers, and health care professionals) and those who *participate* to ensure that eHealth technologies fit in with the needs and values (eg, end users and health care insurers/payers) [110]. Project management is needed to avoid a design-build-run-and-see-what-happens approach. We see development as a cyclic process of ideation, designing, building, and evaluating a technology. Consequently a *multidisciplinary team of researchers and developers* (designers, technicians, health care professionals, and health care researchers) must guide the project management and conduct the *planning* in time and space [48]. Project management also requires logistical planning of how, when, and with what purpose *stakeholders* should participate in the research and development [57,58,110].

Contextual Inquiry

Contextual inquiry entails information gathering from the intended users and the environment in which the technology will be implemented. Field observations and interviews with the intended users are needed to obtain insights into the users' day-to-day rituals and habits and how technology can be matched to that. Through workshops, stakeholders (including users) should be invited to discuss the problems and needs and the goals of the eHealth project via personas and scenarios that represent the goals, tasks, actions, or decisions that are relevant to support the technology [111]. Stakeholders with different backgrounds identify their problems with the current health care delivery via the scenarios and articulate their ideas about how to solve the problems. In addition to this, they define who the relevant stakeholders (key stakeholders with a vested interest)

XSL•FO
RenderX

are who are affected by the problems and solutions. It is important that the opinions of all those involved be taken into account, as exclusion can cause a negative effect on future collaboration. To facilitate the discussion and subsequent reflections, scenarios can be used that present conceptual models and multiangle viewpoints (political, social, clinical, and behavioral) [9,23].

Value Specification

Consequently, *value specification* implies the recognition and quantification of the economic, medical, social, or behavioral values of the key stakeholders [58,110]. The most favorable solutions along with user and organizational requirements emerge from this process (user requirements, value drivers; see Figure 1). The value specification process elaborates on the outcomes of the contextual inquiry. In this cycle, the key stakeholders determine their values (economic, social, and behavioral) and rank them based on the importance of finding solutions for the identified problem(s).

The value-ranking method we use is based on multicriteria decision-making techniques (such as the analytic hierarchy/network process) that score attributes relatively and according to their hierarchy [112]. Value specification refers to goal setting and to defining the functional and organizational requirements to realize the values. It is aimed at exploring what health care improvements are foreseen and what the possibilities or expected limitations are to realize the values. The specified values have to be translated by the stakeholders into functionalities of the design and critical factors for the implementation. For example, during the process of developing a teledermatology application, the key stakeholders identified problems with measuring the possible risks of infection of diabetic feet and insufficient communication between caregivers (general practitioner and dermatologist). The values they formulated were higher quality of care and efficiency to reduce the number of errors and misinterpretations. The technology should therefore have functions to measure the conditions of the wound in an objective and standard way, and the measurements should be communicable in a standardized way. At the same time, the development team gains insights into how to shape the business to offer the values. For example: what are the costs and benefits of teledermatology for the general practitioner, specialist, and patients? The best solution to the problem, the one that emerges from ranking the obtained values with the stakeholders, is the one that will be most beneficial to, and favored by, the key stakeholders.

Design

This is followed by *design*, which refers to building prototypes that fit with the values and user requirements. The design cycle involves the translation of functional requirements into technical requirements and prototypes, given the specified values and goals of the eHealth project. The project management team has to visualize the translation into mock-ups keeping in mind the values, goals, and tasks that have to be fulfilled. Mock-ups, storyboards, or paper prototypes [101] are created and tested sequentially and iteratively with intended users [23,26,29,48,60] and, as a result, the prototypes are refined. The prototypes are tested in real-life situations. The intended users are invited in several rounds via concrete scenarios or tasks to give feedback and to test whether the prototypes match their expectations and mental models (way of thinking and working). To fine-tune the format and content, persuasive techniques [71,73] and card sorting [113,114] can be used to match the information to users' needs. For example, to increase adherence to medical protocols, these documents can be made more user friendly via Web-based communication systems using card sorting to attune the information structure to their mental models and information-searching behavior [21]. In general, the quality of the design can be assessed at different levels [32]: *system quality*, creating technology that is user friendly, is easy to manage, and matches end users' profiles and roles or tasks in the care-delivery process; *content quality*, creating information that is meaningful (accuracy, legibility, comprehensiveness, consistency, and reliability) and persuasive (format fits with users profile); and *service quality*, providing an e-service that is adequate (timely, responsive, and empathetic) and feasible, and measuring the degree to which the e-service is compatible with the health care system.

Operationalization

Operationalization concerns the actual introduction, adoption, and employment of the technology in practice. The cycle consists of enabling and reinforcing activities and mobilizing resources for training, education, and deployment of the eHealth technology in daily practice. Disregarding these conditions may limit the technology's usefulness and delay decision making. An operationalization plan is needed to guide the adoption process—for example, regulations, opinion leaders, triggers, and incentives for using the eHealth technologies [115]—and to create momentum for managing the innovation [28,116]. A business case can be developed that consists of several scenarios, in-depth financial analyses, details about arrangements with other organizations, concrete plans for roles and activities, etc. The implementation of a prototype is discussed via filling up a *business model* canvas [58,106,107] with obtained critical factors, which allows discussion on how to form the business and what strategic choices must be made in order to implement the eHealth technology. A business model is to be developed to steer the adoption process—for example, with regard to internal and external incentives for using the eHealth technologies (the details of business modeling will be explained in a subsequent paper).

Summative Evaluation

Finally, *summative evaluation* refers to the actual uptake of a technology (its usage) and the assessment of the impact of eHealth technologies in terms of clinical, organizational, and behavioral terms. The summative evaluation measures the outcomes at different levels: the usage of a technology and the effects on performance criteria for high-quality care [28,117]. The critical factors that became apparent influence the uptake and impact of the eHealth technology and therefore need to be closely monitored. If certain critical factors start to have negative effects in the summative evaluation phase, the choice needs to be made to iterate to change and improve the current implementation or totally redesign the implementation. This

XSL•FO
RenderX

way, the eHealth technology can be kept sustainable and cost effective.

Conclusion

In this paper we have argued for a holistic approach for the development of eHealth technologies that integrates persuasive health technology theories with a managerial approach (business modeling) to improve the uptake and impact of eHealth technologies in practice. Based on reviews of current eHealth frameworks and on empirical research, we formulated principles for developing eHealth technologies. These principles are the bedrock of the holistic framework we introduced in this paper. The framework provides a comprehensive development strategy; it is suggested that, in the real world, time, policy, and financial considerations can hinder the use of the full framework. The framework is flexible and provides strategies that can be used in part and in a forward (development) and backward (summative evaluation) process. Deficiencies in these processes can be recognized and articulated to determine the bandwidth of innovations. The framework can serve as a debating instrument to clarify areas that would otherwise remain unanswered, unclear, or unknown.

The framework is currently being applied in several research projects. The preliminary results show the benefits of the holistic approach. Technology is not considered as a tool or end in itself, but as a catalyst for innovation. In a teledermatology project about wound care (see Multimedia Appendix 4), it became clear via contextual inquiry that stakeholders demanded a more comprehensive solution than just a tool for taking pictures of a wound. Via value creation, a new infrastructure for replacing hospital care with home-based care was developed with the aid of the roadmap. Via business modeling, the project management team had to think about how technology could improve the wound care process and what the implications were for replacing hospital care with home-based care from a socioeconomic and medical perspective. This resulted in a new infrastructure for teledermatology in health care and a business model that guided the deployment of the eHealth technology. The implications for reimbursement and medical practice were articulated before the production of eHealth technology, and as such the critical factors for deployment were translated into business models that could be discussed with key stakeholders to find out what model suited the eHealth technology best.

Stakeholder engagement resulted in commitment, trust, and a positive attitude toward investments in eHealth technologies (findings from the EurSafety Health-net project; see Multimedia Appendix 4). The participatory development of end users throughout the development process resulted in better adherence to technologies and fewer errors. Moreover, stakeholder involvement resulted in a rethinking of how technology can innovate health care. Standard care tends to be protocol driven; these protocols are often impossible to find or are inaccurate, or too rigid, to be manageable in practice. In the EurSafety Health-net project, medical staff were involved in the codesign of a new approach for antibiotic stewardship (to avoid resistance to infections). The value-creation process resulted in a reconsideration of the values of medical thinking and how technology can fit in with medical practice. This resulted in a shift from solely protocol-driven thinking to an infrastructure for the improved management of antibiotics via Web-based systems for communication and information.

In a companion paper in this journal, we elaborate on the business modeling aspects required to foster the sustainability of eHealth technologies [118]. To support a discussion about the development of eHealth technologies we created a Wiki to accompany our framework. The Wiki, available at http://ehealthwiki.org, is an open and collaborative approach to the development of eHealth technologies. It will provide an expanding and continuously evolving collection of instruments and tools to assist developers, researchers and policy makers. In a following paper we will describe the Wiki.

Acknowledgments

We thank Lex van Velsen, University of Twente, Enschede, The Netherlands, for his cooperation

Conflicts of Interest

None declared.

Editorial note: One of the authors (GE) is also editor of the Journal of Medical Internet Research (JMIR). Because of his involvement in the conduct of this research and writing of this paper, assessment and peer review have been carried out entirely by an associate editor (Potts). GE has not been involved in any editorial decisions related to this paper.

Multimedia Appendix 1

In- and excluded eHealth journal papers.

[PDF File (Adobe PDF File), 36KB - jmir_v13i4e111_app1.pdf]

Multimedia Appendix 2

Target group, goal, foundation, definition of technology.

[PDF File (Adobe PDF File), 152KB - jmir_v13i4e111_app2.pdf]

XSL•FO
RenderX

Multimedia Appendix 3

Strategies and principles for evaluation, design, and implementation of eHts.

[PDF File (Adobe PDF File), 142KB - jmir_v13i4e111_app3.pdf]

Multimedia Appendix 4

Case studies.

[PDF File (Adobe PDF File), 12KB - jmir_v13i4e111_app4.pdf]

References

1. Atienza AA, Hesse BW, Gustafson DH, Croyle RT. E-health research and patient-centered care examining theory, methods, and application. Am J Prev Med 2010 Jan;38(1):85-88. [doi: 10.1016/j.amepre.2009.10.027] [Medline: 20117562]
2. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
3. 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]
4. World Health Organization. Medical Devices: Managing the Mismatch: An Outcome of the Priority Medical Devices Project. Geneva, Switzerland: World Health Organization; 2010.
5. Chaudhry SI, Phillips CO, Stewart SS, Riegel B, Mattera JA, Jerant AF, et al. Telemonitoring for patients with chronic heart failure: a systematic review. J Card Fail 2007 Feb;13(1):56-62. [doi: 10.1016/j.cardfail.2006.09.001] [Medline: 17339004]
6. Dansky KH, Thompson D, Sanner T. A framework for evaluating eHealth research. Eval Program Plann 2006 Nov;29(4):397-404. [doi: 10.1016/j.evalprogplan.2006.08.009] [Medline: 17950868]
7. Kelders SM, van Gemert-Pijnen JE, Werkman A, Seydel ER. Evaluation of a web-based lifestyle coach designed to maintain a healthy bodyweight. J Telemed Telecare 2010;16(1):3-7. [doi: 10.1258/jtt.2009.001003] [Medline: 20086259]
8. Nijland N, Cranen K, Boer H, van Gemert-Pijnen JE, Seydel ER. Patient use and compliance with medical advice delivered by a web-based triage system in primary care. J Telemed Telecare 2010;16(1):8-11. [doi: 10.1258/jtt.2009.001004] [Medline: 20086260]
9. Nijland N, van Gemert-Pijnen J, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]
10. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
11. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006;8(3):e20 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
12. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
13. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
14. Kelders SM, Van Gemert-Pijnen JE, Werkman A, Nijland N, Seydel ER. Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
15. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. J Med Internet Res 2010;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]
16. Van 't Riet J, Crutzen R, De Vries H. Investigating predictors of visiting, using, and revisiting an online health-communication program: a longitudinal study. J Med Internet Res 2010;12(3):e37 [FREE Full text] [doi: 10.2196/jmir.1345] [Medline: 20813716]
17. Wangberg SC, Bergmo TS, Johnsen JA. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
18. Wanner M, Martin-Diener E, Bauer G, Braun-Fahrländer C, Martin BW. Comparison of trial participants and open access users of a web-based physical activity intervention regarding adherence, attrition, and repeated participation. J Med Internet Res 2010;12(1):e3 [FREE Full text] [doi: 10.2196/jmir.1361] [Medline: 20147006]
19. Kelders SM, van Gemert-Pijnen JE, Werkman A, Seydel ER. Usage and effect of a web-based intervention for the prevention of overweight; a RCT. Stud Health Technol Inform 2010;160(Pt 1):28-32. [Medline: 20841644]

XSL•FO
RenderX

20. Nijland N. Grounding eHealth: Towards a Holistic Framework for Sustainable eHealth Technologies [thesis]. Enschede: University of Twente; 2011.
21. Verhoeven F. When Staff Handle Staph: User-Driven Versus Expert-Driven Communication of Infection Control Guidelines [thesis]. Enschede: University of Twente; 2009.
22. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen L. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May;4(3):666-684 [FREE Full text] [Medline: 20513335]
23. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]
24. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
25. Van De Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of Health 2.0 and Medicine 2.0: a systematic review. J Med Internet Res 2010;12(2):e18 [FREE Full text] [doi: 10.2196/jmir.1350] [Medline: 20542857]
26. Kukafka R, Johnson SB, Linfante A, Allegrante JP. Grounding a new information technology implementation framework in behavioral science: a systematic analysis of the literature on IT use. J Biomed Inform 2003 Jun;36(3):218-227. [Medline: 14615230]
27. Shaw NT. 'CHEATS': a generic information communication technology (ICT) evaluation framework. Comput Biol Med 2002 May;32(3):209-220. [Medline: 11922936]
28. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). Int J Med Inform 2008 Jun;77(6):386-398. [doi: 10.1016/j.ijmedinf.2007.08.011] [Medline: 17964851]
29. Kaufman D, Roberts WD, Merrill J, Lai TY, Bakken S. Applying an evaluation framework for health information system design, development, and implementation. Nurs Res 2006;55(2 Suppl):S37-S42. [Medline: 16601633]
30. Van Der Meijden MJ, Tange HJ, Troost J, Hasman A. Determinants of success of inpatient clinical information systems: a literature review. J Am Med Inform Assoc 2003;10(3):235-243 [FREE Full text] [doi: 10.1197/jamia.M1094] [Medline: 12626373]
31. DeLone W, McLean E. Information systems success: the quest for the dependent variable. Inf Syst Res 1992;3(1):60-95. [doi: 10.1287/isre.3.1.60]
32. DeLone W, McLean E. The DeLone and McLean model of information systems success: a ten-year update. J Manage Inform Syst 2003;19(4):9-30.
33. Hebert M. Telehealth success: evaluation framework development. Stud Health Technol Inform 2001;84(Pt 2):1145-1149. [Medline: 11604908]
34. Donabedian A. Evaluating the quality of medical care. Milbank Mem Fund Q 1966 Jul;44(3):Suppl:166-Suppl:206. [Medline: 5338568]
35. National Cancer Institute. Making Health Communication Programs Work. Bethesda, MD: US Department of Health and Human Services; 1989.
36. Boehm BW. A spiral model of software development and enhancement. Computer 1988;21(5):61-72. [doi: 10.1109/2.59]
37. 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]
38. Royce WW. Managing the development of large software systems: concepts and techniques. In: 9th International Conference on Software Engineering (International Conference on Software Engineering//Proceedings).: Association for Computing Machinery; 1987 Presented at: 9th International Conference on Software Engineering; Apr 2, 1987; Los Alamitos, CA, USA.
39. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-340. [doi: 10.2307/249008]
40. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. Int J Man Mach Stud 1993;38(3):475-487. [doi: 10.1006/imms.1993.1022]
41. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478.
42. Miller EA. Telemedicine and doctor-patient communication: a theoretical framework for evaluation. J Telemed Telecare 2002;8(6):311-318. [doi: 10.1258/135763302320939185] [Medline: 12537917]
43. Field ME. Telemedicine: A Guide to Assessing Telecommunications in Health Care. Washington, DC: National Academy Press; 1996.
44. Reid JA. Telemedicine Primer: Understanding the Issues. Billings, MT: Innovative Medical Communications; 1996.
45. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
46. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]

XSL•FO
RenderX

47. Robinson TN, Patrick K, Eng TR, Gustafson D. An evidence-based approach to interactive health communication: a challenge to medicine in the information age. Science Panel on Interactive Communication and Health. JAMA 1998 Oct 14;280(14):1264-1269. [Medline: 9786378]
48. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
49. Kazanjian A, Green CJ. Beyond effectiveness: the evaluation of information systems using A Comprehensive Health Technology Assessment Framework. Comput Biol Med 2002 May;32(3):165-177. [Medline: 11922933]
50. Hamid A, Sarmad A. Evaluation of e-health services: user's perspective criteria. Transform Gov 2008;2(4):243-255. [doi: 10.1108/17506160810917945]
51. Eng TR, Gustafson DH, Henderson J, Jimison H, Patrick K. Introduction to evaluation of interactive health communication applications. Science Panel on Interactive Communication and Health. Am J Prev Med 1999 Jan;16(1):10-15. [Medline: 9894549]
52. Jai Ganesh A. E-health: drivers, applications, challenges ahead and strategies: a conceptual framework. Ind J Med Inform 2004;1:39-47.
53. Honderich T. The Oxford Companion to Philosophy: New Edition. New York, NY: Oxford University Press; 2005.
54. Lawrence C, Weisz G. Greater Than the Parts: Holism in Biomedicine, 1920-1950. New York, NY: Oxford University Press; 1998.
55. James S. The Content of Social Explanation. Cambridge: Cambridge University Press; 1984.
56. Nijland N, van Gemert-Pijnen JE, Boer H, Steehouder MF, Seydel ER. Increasing the use of e-consultation in primary care: results of an online survey among non-users of e-consultation. Int J Med Inform 2009 Oct;78(10):688-703. [doi: 10.1016/j.ijmedinf.2009.06.002] [Medline: 19625210]
57. Carr D, Howells A, Chang M, Hirji N, English A. An integrated approach to stakeholder engagement. Healthc Q 2009;12 Spec No Ontario:62-70 [FREE Full text] [Medline: 19458512]
58. Freeman R. The stakeholder approach revisited. Z Wirtsch Unternehmensethik 2004;5(3):228-241.
59. Hesse BW, Shneiderman B. eHealth research from the user's perspective. Am J Prev Med 2007 May;32(5 Suppl):S97-103. [doi: 10.1016/j.amepre.2007.01.019] [Medline: 17466825]
60. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Comput Biol Med 2002 May;32(3):141-149. [Medline: 11922931]
61. Schön D. The Reflective Practitioner: How Professionals Think in Action. New York, NY: Basic Books; 1983.
62. Lilford RJ, Foster J, Pringle M. Evaluating eHealth: how to make evaluation more methodologically robust. PLoS Med 2009 Nov;6(11):e1000186 [FREE Full text] [doi: 10.1371/journal.pmed.1000186] [Medline: 19956674]
63. Broens TH, Huis in't Veld RM, Vollenbroek-Hutten MM, Hermens HJ, van Halteren AT, Nieuwenhuis LJ. Determinants of successful telemedicine implementations: a literature study. J Telemed Telecare 2007;13(6):303-309. [doi: 10.1258/135763307781644951] [Medline: 17785027]
64. Grigsby J, Brega AG, Devore PA. The evaluation of telemedicine and health services research. Telemed J E Health 2005 Jun;11(3):317-328. [doi: 10.1089/tmj.2005.11.317] [Medline: 16035929]
65. May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, et al. Understanding the implementation of complex interventions in health care: the normalization process model. BMC Health Serv Res 2007;7:148 [FREE Full text] [doi: 10.1186/1472-6963-7-148] [Medline: 17880693]
66. Øvretveit J, Scott T, Rundall TG, Shortell SM, Brommels M. Improving quality through effective implementation of information technology in healthcare. Int J Qual Health Care 2007 Oct;19(5):259-266 [FREE Full text] [doi: 10.1093/intqhc/mzm031] [Medline: 17717038]
67. Shreeve S. Crossover Health. 2007. Health 2. 0: The Definition URL: http://blog.crossoverhealth.com/2007/01/24/health-20-the-definition/ [accessed 2010-09-14] [WebCite Cache ID 5jmdm42UE]
68. Arnrich B, Mayora O, Bardram J, Tröster G. Pervasive healthcare: paving the way for a pervasive, user-centered and preventive healthcare model. Methods Inf Med 2010;49(1):67-73. [doi: 10.3414/ME09-02-0044] [Medline: 20011810]
69. Demiris G, Afrin LB, Speedie S, Courtney KL, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]
70. Flynn D, Gregory P, Makki H, Gabbay M. Expectations and experiences of eHealth in primary care: a qualitative practice-based investigation. Int J Med Inform 2009 Sep;78(9):588-604. [doi: 10.1016/j.ijmedinf.2009.03.008] [Medline: 19482542]
71. Chatterjee S, Price A. Healthy living with persuasive technologies: framework, issues, and challenges. J Am Med Inform Assoc 2009;16(2):171-178 [FREE Full text] [doi: 10.1197/jamia.M2859] [Medline: 19074300]
72. Fogg B. Persuasive Technology: Using Computers to Change What We Think and Do. San Francisco, CA: Morgan Kaufmann; 2002.
73. Fogg B. A behavior model for persuasive design. In: Persuasive '09. 2009 Presented at: 4th International Conference on Persuasive Technology; April 26-29, 2009; Claremont, CA, USA. [doi: 10.1145/1541948.1541999]

XSL•FO
RenderX

74. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(1):28.
75. Ploug T, Hasle P, Oinas-Kukkonen H. Persuasive technology. In: Persuasive Technology: 5th International Conference, PERSUASIVE 2010, Copenhagen, Denmark, June 7-10, 2010, Proceedings (Lecture Notes in Computer ... Applications, incl. Internet/Web, and HCI).: Springer; 2010 Presented at: 5th International Conference on Persuasive Technology; June 7-10, 2010; Copenhagen, Denmark.
76. Ahern DK. Challenges and opportunities of eHealth research. Am J Prev Med 2007 May;32(5 Suppl):S75-S82. [doi: 10.1016/j.amepre.2007.01.016] [Medline: 17466822]
77. Ammenwerth E, Gräber S, Herrmann G, Bürkle T, König J. Evaluation of health information systems-problems and challenges. Int J Med Inform 2003 Sep;71(2-3):125-135. [Medline: 14519405]
78. Atienza AA, Hesse BW, Baker TB, Abrams DB, Rimer BK, Croyle RT, et al. Critical issues in eHealth research. Am J Prev Med 2007 May;32(5 Suppl):S71-S74. [doi: 10.1016/j.amepre.2007.02.013] [Medline: 17466821]
79. Bates DW, Wright A. Evaluating eHealth: undertaking robust international cross-cultural eHealth research. PLoS Med 2009 Sep;6(9):e1000105 [FREE Full text] [doi: 10.1371/journal.pmed.1000105] [Medline: 19753106]
80. Currie LM. Evaluation frameworks for nursing informatics. Int J Med Inform 2005 Dec;74(11-12):908-916. [doi: 10.1016/j.ijmedinf.2005.07.007] [Medline: 16099711]
81. Eng TR. eHealth research and evaluation: challenges and opportunities. J Health Commun 2002;7(4):267-272. [doi: 10.1080/10810730290001747] [Medline: 12356287]
82. 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. [doi: 10.1258/135763305775124722] [Medline: 16375790]
83. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
84. Gustafson DH, Wyatt JC. Evaluation of ehealth systems and services. BMJ 2004 May 15;328(7449):1150 [FREE Full text] [doi: 10.1136/bmj.328.7449.1150] [Medline: 15142895]
85. Rigby M. Essential prerequisites to the safe and effective widespread roll-out of e-working in healthcare. Int J Med Inform 2006 Feb;75(2):138-147. [doi: 10.1016/j.ijmedinf.2005.06.006] [Medline: 16085451]
86. Taylor P. Evaluating telemedicine systems and services. J Telemed Telecare 2005;11(4):167-177. [doi: 10.1258/1357633054068955] [Medline: 15969791]
87. Yusof MM, Papazafeiropoulou A, Paul RJ, Stergioulas LK. Investigating evaluation frameworks for health information systems. Int J Med Inform 2008 Jun;77(6):377-385. [doi: 10.1016/j.ijmedinf.2007.08.004] [Medline: 17904898]
88. Cockerill R, Barnsley J. Innovation theory and its applicability to our understanding of the diffusion of new management practices in health care organizations. Healthc Manage Forum 1997;10(1):35-38. [Medline: 10167073]
89. Kern J. Evaluation of teleconsultation systems. Int J Med Inform 2006;75(3-4):330-334. [doi: 10.1016/j.ijmedinf.2005.08.001] [Medline: 16143564]
90. Wootton R, Hebert MA. What constitutes success in telehealth? J Telemed Telecare 2001;7 Suppl 2:3-7. [Medline: 11747643]
91. May C, Harrison R, Finch T, MacFarlane A, Mair F, Wallace P, Telemedicine Adoption Study Group. Understanding the normalization of telemedicine services through qualitative evaluation. J Am Med Inform Assoc 2003;10(6):596-604 [FREE Full text] [doi: 10.1197/jamia.M1145] [Medline: 12925553]
92. 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]
93. Resnicow K, Strecher V, Couper M, Chua H, Little R, Nair V, et al. Methodologic and design issues in patient-centered e-health research. Am J Prev Med 2010 Jan;38(1):98-102. [doi: 10.1016/j.amepre.2009.09.034] [Medline: 20117564]
94. Kaplan B. Evaluating informatics applications--clinical decision support systems literature review. Int J Med Inform 2001 Nov;64(1):15-37. [Medline: 11673100]
95. Creswell JW, Plano Clark VL, Guttman M, Hanson W. Advanced mixed methods research designs. In: Tashakkori A, Teddlie C, editors. Handbook of Mixed Methods in Social & Behavioral Research. Thousand Oaks, CA: Sage Publications; 2003:209-240.
96. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2010 Mar 30. [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
97. Gagnon MP, Légaré F, Labrecque M, Frémont P, Pluye P, Gagnon J, et al. Interventions for promoting information and communication technologies adoption in healthcare professionals. Cochrane Database Syst Rev 2009(1):CD006093. [doi: 10.1002/14651858.CD006093.pub2] [Medline: 19160265]
98. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
99. Verhoeven F, van Gemert-Pijnen L, Dijkstra K, Nijland N, Seydel E, Steehouder M. The contribution of teleconsultation and videoconferencing to diabetes care: a systematic literature review. J Med Internet Res 2007;9(5):e37 [FREE Full text] [doi: 10.2196/jmir.9.5.e37] [Medline: 18093904]

XSL•FO
RenderX

100. Demiris G, Charness N, Krupinski E, Ben-Arieh D, Washington K, Wu J, et al. The role of human factors in telehealth. Telemed J E Health 2010 May;16(4):446-453. [doi: 10.1089/tmj.2009.0114] [Medline: 20420540]
101. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
102. Zhang J. Human-centered computing in health information systems. Part 1: analysis and design. J Biomed Inform 2005 Feb;38(1):1-3. [doi: 10.1016/j.jbi.2004.12.002] [Medline: 15694880]
103. Zhang J. Human-centered computing in health information systems part 2: evaluation. J Biomed Inform 2005 Jun;38(3):173-175. [doi: 10.1016/j.jbi.2004.12.005] [Medline: 15896690]
104. Esser PE, Goossens RH. A framework for the design of user-centred teleconsulting systems. J Telemed Telecare 2009;15(1):32-39. [doi: 10.1258/jtt.2008.080601] [Medline: 19139218]
105. Curry SJ. eHealth research and healthcare delivery beyond intervention effectiveness. Am J Prev Med 2007 May;32(5 Suppl):S127-S130. [doi: 10.1016/j.amepre.2007.01.026] [Medline: 17466817]
106. Chesbrough H. Open Business Models: How to Thrive in the New Innovation Landscape. Boston, MA: Harvard Business School Press; 2006.
107. Osterwalder A, Pigneur Y. Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. Hoboken, NJ: John Wiley & Sons; 2010.
108. Osterwalder A, Pigneur Y, Tucci C. Clarifying business models: origins, present, and future of the concept. Commun Assoc Inf Syst 2005;16(1):1-25.
109. Parente ST. Beyond the hype: a taxonomy of e-health business models. Health Aff (Millwood) 2000;19(6):89-102 [FREE Full text] [Medline: 11192424]
110. Sharp H, Finkelstein A, Galal G. Stakeholder identification in the requirements engineering process. In: DEXA '99. 1999 Presented at: 10th International Conference and Workshop on Database and Expert Systems Applications; Sept 1-3, 1999; Florence, Italy. [doi: 10.1109/DEXA.1999.795198]
111. Lerouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inform 2011 Apr 8. [doi: 10.1016/j.ijmedinf.2011.03.006] [Medline: 21481635]
112. Saaty TL. Relative measurement and its generalization in decision making: why pairwise comparisons are central in mathematics for the measurement of intangible factors - the analytic hierarchy/network process. Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas 2008;102(2):251-318. [doi: 10.1007/BF03191825]
113. Spencer D. Card Sorting: Designing Usable Categories. Brooklyn, NY: Rosenfeld Media; 2009.
114. Wood JR, Wood LE. Card sorting: current practices and beyond. J Usability Stud 2008;4(1):1-6.
115. Cain M, Mittman R. Diffusion of Innovation in Health Care. Oakland, CA: California HealthCare Foundation; 2002.
116. Tidd J. Gaining Momentum: Managing the Diffusion of Innovations. London: Imperial College Press; 2010.
117. Eysenbach G. A framework for evaluating e-health: systematic review of studies assessing the quality of health information and services for patients on the Internet [abstract]. J Med Internet Res 2000;2(suppl 2). [doi: 10.2196/jmir.2.suppl2.e13]
118. Van Limburg M, Van Gemert-Pijnen JE, Nijland N, Ossebaard HC, Hendrix RMG, Seydel ER. Why business modelling is crucial in the development of eHealth technologies. J Med Internet Res (forthcoming) 2011;13(4). [doi: 10.2196/jmir.1674]

Abbreviations

fr.: framework
RCT: randomized controlled trial

Edited by H Potts; submitted 20.10.10; peer-reviewed by P Esser, C Vincent, WAR Wan Mohd Isa; comments to author 06.01.11; revised version received 20.07.11; accepted 31.08.11; published 13.12.11.

Please cite as:
van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, Seydel ER
A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies
J Med Internet Res 2011;13(4):e111
URL: http://www.jmir.org/2011/4/e111/
doi:10.2196/jmir.1672
PMID:22155738

©Julia EWC van Gemert-Pijnen, Nicol Nijland, Maarten van Limburg, Hans C Ossebaard, Saskia M Kelders, Gunther Eysenbach, Erwin R Seydel. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 13.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Why Business Modeling is Crucial in the Development of eHealth Technologies

Maarten van Limburg[1], MSc BEng; Julia EWC van Gemert-Pijnen[1], PhD; Nicol Nijland[1], PhD; Hans C Ossebaard[2], MA; Ron MG Hendrix[1], MD, PhD; Erwin R Seydel[1], PhD

[1]Center for eHealth Research and Disease Management, Department of Psychology, Health and Technology, Faculty of Behavioural Sciences, University of Twente, Enschede, Netherlands

[2]National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands

Corresponding Author:
Maarten van Limburg, MSc BEng
Center for eHealth Research and Disease Management
Department of Psychology, Health and Technology, Faculty of Behavioural Sciences
University of Twente
Drienerlolaan 5
Enschede, 7500AE
Netherlands
Phone: 31 534896070
Fax: 31 534892388
Email: a.h.m.vanlimburg@utwente.nl

Abstract

The impact and uptake of information and communication technologies that support health care are rather low. Current frameworks for eHealth development suffer from a lack of fitting infrastructures, inability to find funding, complications with scalability, and uncertainties regarding effectiveness and sustainability. These issues can be addressed by defining a better implementation strategy early in the development of eHealth technologies. A business model, and thus business modeling, help to determine such an implementation strategy by involving all important stakeholders in a value-driven dialogue on what the technology should accomplish. This idea also seems promising to eHealth, as it can contribute to the whole development of eHealth technology. We therefore suggest that business modeling can be used as an effective approach to supporting holistic development of eHealth technologies. The contribution of business modeling is elaborated in this paper through a literature review that covers the latest business model research, concepts from the latest eHealth and persuasive technology research, evaluation and insights from our prior eHealth research, as well as the review conducted in the first paper of this series. Business modeling focuses on generating a collaborative effort of value cocreation in which all stakeholders reflect on the value needs of the others. The resulting business model acts as the basis for implementation. The development of eHealth technology should focus more on the context by emphasizing what this technology should contribute in practice to the needs of all involved stakeholders. Incorporating the idea of business modeling helps to cocreate and formulate a set of critical success factors that will influence the sustainability and effectiveness of eHealth technology.

(J Med Internet Res 2011;13(4):e124) doi:10.2196/jmir.1674

KEYWORDS

Business model; cocreation; collaboration; eHealth; implementation; multidisciplinary; stakeholder; sustainability; value creation

Introduction

Health care systems worldwide will face sustainability problems in the near future caused by a tension between an increasing demand for and a mismatch in the supply of health care services [1]. The growing demand for health care services is generally explained by an aging population and the rise in prevalence and incidence of chronic diseases and obesity. In addition, these increased demands imply increased complexity of treatments due to rapid advances in medical technology and increased comorbidity [1,2]. At the same time, the health care industry struggles with inefficiencies in procurement of supplies and inadequate use or lack of resources. In the United States, for example, the financial consequences of inefficiency are estimated to be in the range of 30% to 40% of total health care costs [3]. Without rapid action, health care services shall soon

XSL•FO
RenderX

become less accessible and unaffordable and will deteriorate in quality.

In many industries, Web-based and mobile technologies have changed and are still changing conventional business activities to Internet-based activities such as Web 2.0 services or e-business [4,5]. In the health care industry, similar opportunities, often called eHealth, seem promising to help solve the aforementioned demand and supply problems in healthcare [6,7]. Indeed, eHealth technologies can contribute to improved communication and information sharing among health professionals, patients, and researchers and aim to improve quality and effectiveness of health care services [6,8,9]. However, eHealth technologies suffer from a range of recurring problems [3,10-16] as outlined in Textbox 1.

These problems can be attributed to insufficient attention to the development process and implementation of eHealth technologies. We believe that in order to tackle the aforementioned problems and to ensure a proper uptake, long-term sustainability, and effectiveness, new development frameworks are needed that make implementation an integral part of eHealth development. We see that implementation of eHealth technologies in practice is underestimated and overlooked in eHealth development approaches. Therefore, we proposed a new holistic approach in our paper, "A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies" [17], which describes the entire development and is aimed at creating a fit between technology, humans, and organizations.

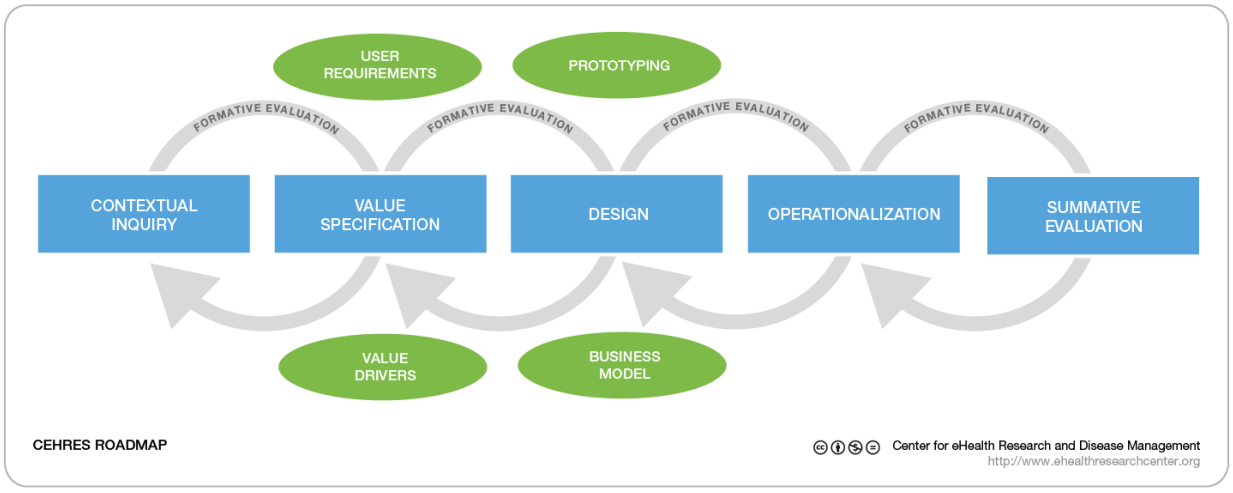
Textbox 1. Recurring Problems of Ehealth Technologies

- Currently established financial structures slow down innovation.
- Necessary legislations for modernizing health care lag behind.
- Involved parties are reluctant and uptake remains low.
- eHealth development focuses too strongly on engineering-driven solutions.
- eHealth technologies are deployed in a fragmented fashion and have poor scalability.
- The number of stakeholders and dependencies cause complexity.
- There is a lack of cost-effectiveness studies.
- eHealth research tends to focus on finding clinical evidence in terms of health outcomes, for example, yet the impact of eHealth technology does not rely solely on clinical evidence; there are more factors that determine the success of eHealth technology.

CeHRes Roadmap

The Center for eHealth Research and Disease Management (CeHRes) Roadmap (Figure 1), introduced in "A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies" in this issue of the *Journal of Medical Internet Research* [17], offers a holistic approach to eHealth development. This roadmap guides the development of persuasive technology and business modeling as interwoven activities. This approach allows eHealth technologies to be designed according to the needs of its users and to fit with their behavior, but also, due to business modeling, it allows the development process to be value-driven. Stakeholders are involved in the development process and, based on their values, an eHealth technology can be designed matching with intended collaboration and cocreation, and eventually an implementation can be found.

Figure 1. CeHRes Roadmap.



XSL•FO
RenderX

Why eHealth Needs Business Modeling

In this paper we focus on business modeling and why it supports the development of eHealth technologies. Business modeling is interwoven with development to make both design and implementation value-driven. After all, it is futile to develop an eHealth technology that does not catch on because in practice it does not match demands or its intended purpose.

Implementation must ensure that an eHealth technology will live up to its fullest potential in real-world conditions and circumstances. In order for eHealth technology to succeed, all organizations have to collaborate and interact, and some organizations have to maintain and perhaps fund the project. eHealth technology needs to fit in existing care infrastructures or, perhaps even more importantly, be a catalyst for new, innovative care infrastructures. In other words, eHealth development encompasses more than technical design. It requires additional research to determine an implementation strategy, that is, a plan to embed technology in its intended practice. Implementation starts with detecting and involving concerned parties and results in a business model that describes the value creation and acts as the basis for a care infrastructure for collaboration and cocreation, possibly with multiple organizations involved. To our knowledge, very few implementation rationales relating to eHealth technologies have been explained. Many of these eHealth technologies are developed with a "jump on the eHealth bandwagon" mentality without clear predetermined goals. Once an eHealth technology has been developed and it becomes apparent that goals are needed, the organization finally starts to think about an implementation strategy. So, current eHealth implementations are usually done *post* development rather than integrated *in* the development process.

Attention to implementation appears too late in the development, and we therefore point out that it is crucial to start preparing an implementation strategy early on. It is better to invest more time and money in researching how eHealth technology can be implemented in its intended care practice than to invest money in an eHealth technology that will not have a satisfying uptake. It happens too often that as soon as research funding stops, an eHealth technology cannot be implemented sustainably, mainly because there is neither support nor interest from other parties. Through business modeling, development of eHealth technologies can be guided with a value-driven evaluation of what is necessary and what is not. Often eHealth technologies are built as replacements for or copies of existing care services and are then fine-tuned for user requirements using user- or human-centered design principles. It is yet to be questioned whether this approach is effective and whether the choices made are really grounded. Business modeling introduces research activities *before* the start of the actual technical design that focus on the context of eHealth technology and provide value drivers that will ground choices of what to develop.

Starting With a Context

An important early step in the development of eHealth technology is analyzing the relevant problem, that is, an eHealth technology is meant to improve a problem of inefficiency or a lack of information or communication. In order to take proper action, the situation needs to be carefully assessed: this is known as *sensemaking* [18]. It is tempting, however, to rush toward thinking of technical solutions for a problem. Such fast solutions may lead to a solution that is technically state-of-the-art but poorly suited to the problem. By analyzing the problem at hand, eHealth technology will gain more context, and this increased understanding will contribute to all further choices that are required in the development process and the implementation. This is why the *contextual inquiry* in our business modeling approach is a crucial first step.

By discussing the problems with all concerned parties (so-called stakeholders, see next paragraph), it becomes clearer which parties will play an important role in the development process and which parties may come to play a role in the implementation of the eHealth technology. Also, this problem-oriented dialogue helps to make these parties more aware of each other's problems, as health care organizations often have limited knowledge of the processes and/or problems that go on at other organizations. In fact, during several of our workshops, it became apparent that people even within the same organization were unaware of each other's exact responsibilities and duties (see Textbox 2 as example).

Textbox 2. Example Case: Finding the Problems With Antibiotics Prescription

Our intention was to understand and improve the behavior behind antibiotics prescription as part of the contextual inquiry for an eHealth technology that is in development. Based on a literature review and expert interviews, we identified the general problems with imprudent antibiotics prescription (causing a high risk of infections), the general prescription process, as well as key stakeholders. We organized a workshop with these key stakeholders within the first hospital ward where we had aimed to start our pilot. These key stakeholders discussed the problems they face daily based on patient scenarios validated by infection experts. This workshop not only enlightened the project management (us) to what problems and opportunities there were, but also created awareness among stakeholders as to what problems *other* stakeholders face and how the mutual problems also affected others. This awareness is vital for the collaboration of these key stakeholders and their future commitment to the project.

Stakeholder Participation

Everyone who affects or is affected by a project is considered a stakeholder [19]. It is therefore critical for the success of eHealth technology to understand the value needs of each stakeholder [20]. Through participation of stakeholders in the development process of eHealth technologies, value needs can be retrieved and a mutually determined fit can be found. According to Pagliari, developing eHealth technologies is a multidisciplinary process [21]. Business modeling deepens this multidisciplinary development of eHealth as it brings multiple stakeholders together in the discussion of the necessary implementation. Business modeling also allows for an

XSL•FO
RenderX

exploration of the value needs of stakeholders that determines both the design of the technology as well as the implementation.

There are many types of stakeholders associated with eHealth: patients, policymakers, vendors, insurers, health care organizations and providers, home care workers, and employers [22]. Therefore, every eHealth technology will have its unique stakeholder network (sometimes also referred to as an ecosystem) that determines potential customer segments and the infrastructure required for value cocreation for eHealth technology. Patients are often overlooked as stakeholders, yet they also have to participate in eHealth development. Patients often use or are subjected to the technology and have legal and social rights to be part of the development [8]. Patient empowerment does not stop at letting patients use eHealth technology; patients should be invited to participate in the development process of technology as well.

The level of engagement determines the salience of each stakeholder to the stakeholder network [23]. In our roadmap, we start by mapping the stakeholder network as part of the contextual inquiry process. As suggested by Sharp, it is best to start with baseline stakeholders (in our approach we start with project initiators) and let them suggest more stakeholders that may be relevant to the eHealth project [24] (see Textbox 3 as an example). We base stakeholder salience on three variables: power, legitimacy, and the urgency of the stakeholder [25]. There are various ways to assess salience. This can be done either by asking experts to score the above variables or by asking the stakeholders to score each other. The next step is to start discussing value with stakeholders. The most salient stakeholders will eventually have a bigger influence on the value drivers than less salient ones.

Textbox 3. Example Case: Finding Stakeholders Through Experts and by "snowballing"

In the early phases of any project, there are one or more initiators involved that can provide a list of baseline stakeholders. In one project, for example, a health information technology (IT) company wanted to develop a personal health record service. We spoke to several opinion leaders in health insurance, eHealth, and patient empowerment to form a stakeholder map specific for the Dutch health care system. In the interviews that followed, these stakeholders also provided more potential stakeholders that were relevant for the project, and so a specific stakeholder map appeared. Later on, this stakeholder map was used to report several business model opportunities to the management of the health IT company.

Cocreation

Cocreation in eHealth has already been introduced in disease management, for example, to streamline health care activities among multiple health care organizations. It also plays a role in patient empowerment, as patients are actively involved in their care [12]. Introduction of eHealth technology is often top-down, that is, technology is mainly determined by management. Obviously, management has an important say in whether or not a technology should be introduced, but in our view, a bottom-up approach is needed as well. This bottom-up approach can mean, for example, that a few specialists from a hospital ward also supply input on how they see technology adding value to their work. This is *value specification* that looks further than human-centered design, as it does not only look at the usability of the technology but much wider, that is, at the intended purpose of the technology and its fit in practice.

Participation of stakeholders in development also involves a political element, in that stakeholders feel they really contribute to the technology, and therefore, they feel more involved and positive toward it than when they are excluded. Dialogue is very important in cocreation [26]. Also, scalability problems can be tackled with business modeling by planning ahead through involving future stakeholders, particularly political or influential stakeholders, early in development to avoid eHealth technology becoming too localized and too narrowly focused.

Cocreation and dialogue with stakeholders requires a willingness to be open with each other. Openness is a way of thinking that is rooted in the opportunities of open source software and Web 2.0 that advocates operating with open systems for mutual benefits and transparency [5]. The *open business model*, as described by Chesbrough, combines this idea of openness with business models and promotes that organizations can embed cocreation and collaboration in their business models for shared benefits [27]. Classic success stories of open business models are the Philips Senseo coffee machine or the budget airline Ryanair. In the eHealth context, open systems are emerging too, such as interoperable electronic health records. Business modeling also pursues openness as multiple organizations cocreate value of technology and share benefits.

Regardless of the industry, traditional boundaries between organizations are becoming fuzzier and open business models pave the way for future collaborative success.

When cocreation is a goal, it will mean that eHealth technologies will be more intricate than one single organization carrying full responsibility, and it will require cooperation of multiple health care organizations. Interorganizational dependencies can be very complex, so exploring benefits and value needs is a complex task that requires input from all involved stakeholders. To cooperate and balance these value needs, health care organizations need to extend beyond their traditional boundaries. This implies a different view of the development process of eHealth technology as well: it is not only an "apparatus" that is being created; there is a whole new underlying infrastructure for collaboration that has to be created as well (see Textbox 4 as example).

Eysenbach [8] observes that social networks, collaboration, and active participation are key elements in today's eHealth. When the opportunities of Web 2.0 technology are used for this collaboration in eHealth, this is often called Health 2.0 or Medicine 2.0. For cocreation and collaboration, an infrastructure such as a social network of organizations is needed as well [26]. Within this infrastructure, stakeholders have to interact to cocreate value to eHealth technology. The stakeholder network that appears in the development process is also the basis for an infrastructure and will eventually become an infrastructure required for the collaboration and cocreation supporting the eHealth technology. This cocreation and collaboration is

XSL•FO
RenderX

ongoing; therefore, it is imperative stakeholders all stay involved and interested in supporting and further developing the technology.

Textbox 4. Example Case: a Service Model for Teledermatology

In a teledermatology project, it became apparent that the stakeholders required more than just a technology for a fitting teledermatology solution, they also required a new infrastructure for a service delivery that, for example, would replace hospital care with home care. Via stakeholder meetings, the possibilities were identified, and scenarios were made that would allow cocreation and collaboration with third parties to implement the technology in practice. This resulted in a service model that described value cocreation between the engineers of the technology company and several health service companies, which was quite different to what the management initially had in mind during the early stages of the project.

Value Drivers in eHealth

Chesbrough emphasizes the importance of an implementation by stating that "a mediocre product with a good business model yields more value than a good product with a mediocre business model" [27]. So, business modeling is crucial for the success of an eHealth technology. Through business modeling, the entire development becomes stakeholder-focused and value-driven. Stakeholders are asked early on what value drivers they expect regarding eHealth technology. These value drivers are relevant for both the design of technology as well as the *design* of the implementation strategy that will determine effectiveness and sustainability of eHealth technology.

Business modeling is a value-driven process and, as such, it is not simply a business model but an extensive process through which early opportunities for an eHealth technology are explored, assessment is made of what is required, a case-specific business model is developed, and the said technology is accordingly implemented. As part of the roadmap, we stress that development is a continuum and thus requires ongoing research activities that include design, evaluation, and redesign. Making a choice based on facts today can be improper a week later when new facts emerge. Web technology in particular is notorious for being relentlessly progressive; thus, adaptability is crucial. Over time, stakeholders can come and go or their value needs change, and the implementation needs to be reevaluated and redesigned. In terms of business models, this is called *business model erosion* [28], and due to this erosion, eHealth technology will be less sustainable and effective. So we need more sustainable methods to ground the eHealth development process and, for this, stakeholders need to be continuously involved in the development process and have their say in an implementation.

Our current approach to business modeling is to hold various workshops with relevant stakeholders to determine problems and opportunities in health care, which role technology can play, and which stakeholders are involved and what their importance is to the developed eHealth technology. Stakeholders at the workshops determine the role that the technology needs to fulfill in practice by forming an infrastructure and also determine what makes or breaks effectiveness and sustainability. All these elements are captured with a business model that can be detailed in a business case for further operationalization and deployment of the eHealth technology.

Value creation is central to business modeling. Obviously, in for-profit contexts, this value is mostly monetary, but other kinds of value drivers can be important too. Especially in the health care context, we often see extra attention paid to nonmonetary values, as health care is a special market. Intel's health care information technology (HIT) value model breaks down value into three levels: monetary value, quantifiable value, and benefits, the latter being, for example, social value or certain qualitative values that are considered beneficial but are hard to express in concrete figures [29]. In our business modeling approach, value drivers can be seen very broadly, that is, anything that a stakeholder considers critical to technology is a relevant value driver worthwhile to research. These values drivers form the basis for the development process and implementation.

Business modeling promotes a value-driven dialogue and promotes better understanding of what should be accomplished with eHealth technology [30]. This value-driven approach allows stakeholders in eHealth technologies to better discuss and reflect on the intended value that technology has to offer to the health care setting. Value drivers can also be initially counterproductive, as, for instance, when a certain stakeholder loses money or influence, this stakeholder will then criticize the technology. These negative value drivers then must be compensated for elsewhere. Also, by determining the overall expected value before designing begins, the assessment will be more profound whether or not eHealth technology is worth the investment. Nevertheless, *value* and *value drivers* remain complex concepts. During the value specification, many values will appear and many will also conflict; hence, dialogue is very important. It can be an extensive task to assess and to clarify to stakeholders what value eHealth technology can create, but without looking into value drivers, exact gains of eHealth investments remain unclear *a priori*, and it will be impossible to find a fitting implementation.

With business modeling, we aggregate all value needs *bottom-up* from the stakeholders, and, through dialogue, we try to cocreate a fit between all the values that will become the overall expected value of the eHealth technology. Value becomes the focal point for technical design and also for the critical success factors [31] required for implementation. In our workshops, we use custom mapping software, to elicit these values from stakeholders and to rank scores to their importance according to the stakeholders. This ranking acts as a way to quantify and prioritize values. (A common method for this is called the analytic hierarchy process [32] that, in short, alters the initial scores given to the values by taking the hierarchy of these values into consideration.) These values are input for the design of an eHealth technology and are the basis for implementation. For example, if the value *security* is given a high score by multiple stakeholders, then during implementation, all security-related choices (eg, collaboration with a good software security company) need to be given serious consideration; otherwise, certain stakeholders

XSL•FO
RenderX

will not consider the technology valuable. This determination also influences the technology itself, that is, security-based features are apparently important, and thus designers and developers should thoroughly research what the security requirements are. Textbox 5 provides another example.

Textbox 5. Example Case: How Value Drivers Can Influence Technical Design

During the problem analysis in the teledermatology project, it was found that there were many additional problems in the whole teledermatology process that the initial design of device did not reflect. In general, the device had to offer support regarding how health care professionals in home care can take pictures of wounds so that wounds can be better diagnosed. Consensus arose among stakeholders that it was necessary to provide standardized guidelines for using the technology. We determined what value drivers were relevant to these guidelines, as without these standardized guidelines, the device would be less useful and thus less valuable to the stakeholders. This process also resulted in technical design additions.

Business Models

As the term *business modeling* implies, its core output is a business model. A business model plays an important part in implementation: it acts as the basis for discussion of value drivers with stakeholders and becomes the basis for further operationalization where the business model is made more concrete through a business case, and, subsequently, the actual deployment of eHealth technology can happen.

Research in business models is relatively new, and, thus far, the term *business model* is still ambiguous in science and in practice [30]. Business models are quite often confused with *business process models* that are used on an operational level to describe detailed operational processes [33]. Also, some people associate business models with detailed financial prognoses, which are actually more characteristic of a *business case*. Osterwalder [34] defines a business model as "the rationale of how an organization creates, delivers, and captures value." By this definition, business models act on a strategic level and can be the basis for more detailed business process models and business cases [35]. In our view, one needs to decide on a business model first in order to develop a business case. The business model can be created early on in the development process. The business case can gradually take shape and the details can be developed while the technology is being designed. Obviously, during the development process, a business model can also be refined or altered depending on unforeseen changes or new insights.

Business models became prominent in the late 1990s when the methods of doing business rapidly grew more complex and interdependent [36]. During that time, Internet-based activities became important assets in value creation and opened possibilities for new moneymaking activities and sped up globalization. Organizations had to change their existing strategies and develop new strategies. Yet, in order to achieve this transformation, organizations required something to plan ahead. This is when the term *business model* became widely adopted. A business model helps to relate all strategically defined critical success factors (critical elements in the achievement of successful value creation) into a working whole [37,38]. As such, they allow managers to understand, communicate, and evaluate the strategy for value creation and to conceptualize the strategy in a concise, modeled form [37]. In this period, numerous new business models emerged, and, coinciding with the popularity of the Internet, these were, in particular, business models that explored the potential of Web 2.0 [4].

A framework that is currently popular for defining a business model is the business model canvas by Osterwalder (depicted in Figure 2) [34]. It describes the whole rationale in nine building blocks. In the middle block is the value proposition, the eHealth technology in this case. The top three blocks on the left-hand side of the diagram deal with the required organizational aspects, that is, the key activities, resources, and partners. The top three blocks on the right-hand side deal with who the customers/users are and how to interact with them. At the bottom are the financial aspects. Creating and offering value generate costs, and a revenue model is necessary to capture value back to at least cover these costs. This canvas is an empty framework or blueprint that can be filled with critical success factors and choices to describe the implementation of an eHealth technology. The framework is useful as it describes the entire value creation logic and is a guide for making sure that all nine aspects necessary for value creation are addressed. The framework also helps to classify and group the components of a business model.

However, the process behind filling this canvas determines the quality of the business model. In Osterwalder's book, *Business Model Generation* [34], a strong focus is on *ideation*, that is, thinking up innovative business models on a very high level of abstraction early on for new businesses. But the canvas can also be filled with value drivers based on the value specification that we apply in our business modeling approach. The chosen, important value drivers from the value specification become critical success factors, as they will determine the success of the implementation of the eHealth technology. We place these in the canvas to get an overview as well as to check if all building blocks received adequate attention from the stakeholders and/or researchers. It is also possible that multiple business models can be formed based on the value drivers gathered from the stakeholders, as the example in Textbox 6 demonstrates.

XSL•FO
RenderX

Textbox 6. Example Case: Multiple Business Model Opportunities for Different Scenarios

The aforementioned service model in the teledermatology example (Textbox 4) resulted in multiple possible business models with different service paths. These were:

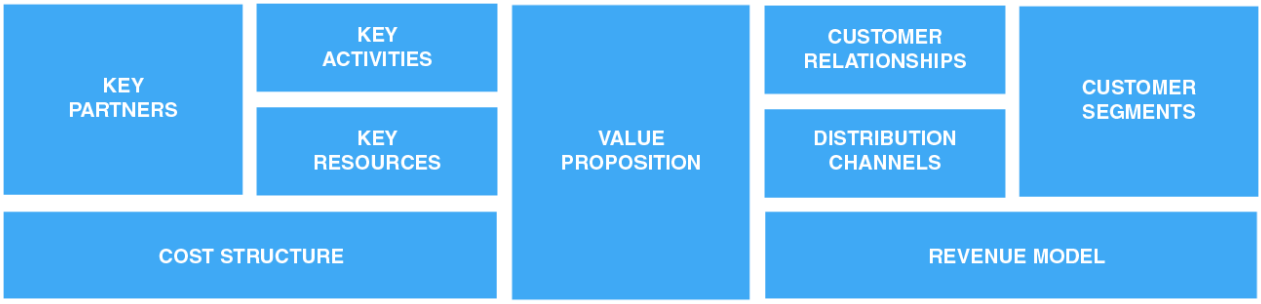
· Keeping everything in-house

· Cocreation with third party organizations that would take care of the teledermatology infrastructure so that the technology company could focus on the technology

· A mix between providing a technology to third parties yet also providing additional technical services to third party organizations in return for a payment for each use

Each business model had its pros and cons, and it was up to the management to decide which of these models they found best fitting to the future of their company.

Figure 2. Business model canvas.



Business Case

Having a business model alone is not enough. Once the desired business model is decided on and all stakeholders agree on the plans, the *operationalization* can be further determined by making a concrete business case based on the business model. A business case contains much more concrete information about the details of the implementation than a business model, but a business model is required to provide an idea of what the implementation should look like. In the business case, concrete descriptions of the necessary activities, resources, and costs can be written down. Usually business cases contain several financial prognoses based on estimated usage of the technology. These prognoses are based on multiple usage scenarios (low, projected, high usage) to gain a better understanding of the dynamics of the costs and potential revenues. Textbox 7 demonstrates how a business case can be made early in a project to demonstrate financial benefits of an eHealth technology. Usually, a business case is continuously updated during the development.

Also in this stage, the required infrastructure that resulted from the stakeholder network and value specification can be further arranged more formally with contracts, formal agreements, and so forth.

Once these steps are taken and the technology is designed, it can be implemented in practice. However, the operationalization is not an endpoint; evaluation is necessary to track whether the technology and implementation still meet the intended goals and whether redesign iterations are necessary in the development.

Textbox 7. Example Case: Business Case for Implementing an Antibiotic Stewardship Program

Changing antibiotic prescription can be beneficial. For example, patients can have a shorter length-of-stay or the prescriber can choose a quicker swap from intravenous to oral antibiotics. Through a calculation, we showed the hospital management that they could save up to a million euros a year on antibiotic costs alone. These financial prognoses convinced the management to start a pilot project for an antibiotic stewardship program.

Evaluation

Development of an eHealth technology starts with a variety of assumptions defined by time or budget constraints. Not everything in a business model can be understood *ab ovo* and requires reflection and progressive insight [39]. By spending more time investigating the exact value needs—even during usage of a technology—the technology and its implementation can be continually refined. As with any technology, eHealth technologies are subject to environmental and contextual changes. Technology never stands still, and most technologies are developed using iterative design approaches [21]. Just as technologies evolve over time, business models are also not static objects [40]. Therefore, *summative and formative evaluation* cannot be performed in an inert state but should be an action or a process (see Textbox 8). Business modeling makes sure technology and implementation keep reflecting on the current and future needs of the stakeholders for sustainability. It is imperative that an eHealth technology remains an object of study even after the technology has been implemented into practice; eHealth technology is not a “fire-and-forget” technology. The evaluation of its success needs to continue for further improvement and anticipation of changes in the health care environment. As a value-driven approach can project the critical success factors, the intended goals of the eHealth technology can be measured.

XSL•FO
RenderX

Textbox 8. Example Case: Summative Evaluation of Web-Based Infection Control System for Methicillin-Resistant Staphylococcus Aureus (mrsa)

In 2008, we launched a website that informs general audience and health care professionals about methicillin-resistant staphylococcus aureus (MRSA). With server logs, we analyzed how the website has been used by visitors over the years and discovered that the chosen card-sort presentation of questions and answers, codesigned in 2008 with the intended users, was indeed effective and could be maintained. Additionally, we found a few ideas for improvements such as improving the search engine optimization, as the number of visitors via Google was significantly growing over the years.

Conclusion

Many eHealth technologies still fail in practice, and little or late attention is given to implementation. We believe preparing the implementation strategy is part of the development process and should start as early as possible in the development. In strategic management, business models are used to define the rationale behind value creation in terms of eHealth, which means the required rationale for implementing an eHealth technology in its care setting. We introduced business modeling as a vital part of our holistic approach for eHealth development in order to improve the uptake and sustainability of eHealth technologies. Business modeling, and our CeHRes Roadmap generally, have proven in multiple, different eHealth projects to be worthwhile in the development of eHealth technologies, helping us to find a better fit among humans, organizations, and technology with a value-driven and stakeholder-focused eHealth development. Business modeling fosters a ground for dialogue regarding the perceived value and purpose of an eHealth technology. An eHealth technology simply has a plethora of stakeholders and they all influence or are influenced by the eHealth technology. Implementation of eHealth technology depends on how well the value needs of stakeholders are met and how they partake in the infrastructure needed for the eHealth technology. Business modeling is a continual activity because the environmental conditions in eHealth are dynamic, so iterative development and anticipation to changes are important for sustainability and long-term success of the technology.

Health care organizations base their operations on century-old reimbursement business models [3]. Progress in medical and technological possibilities and many sociopolitical factors have altered the processes but left settled business models unchanged. Lagging legislation, financial complexity, and a status quo of roles and dependencies seem only to work in favor of perpetuating these inefficient health care processes. Evidential benefits from eHealth technologies remain unsure, as new technological possibilities often cause extra side processes rather than an efficient replacement for the processes that need to be improved. eHealth should not be an irrelevant remake of old processes. Innovative eHealth business models require that core conceptions, current roles, and processes are reevaluated and overhauled from complex organization-centered health care chains to efficient patient-centered health care networks in which multiple health care organizations collaborate to provide care.

eHealth projects need to research new business models. Both in practice as in academic context, a business model is often mentioned as a kind of panacea to improve the effectiveness and sustainability of eHealth technologies; however, the exact *why* and *how* are omitted from the arguments. Often generic business models from other industries (at the so-called taxonomy level) are mentioned as potential solutions which are *per se* unsuited, for example, taxonomies such as subscription-business models or pay-per-click-business models. These generic business models are excellent for classification, but for implementing an eHealth technology, this level-of-detail will not suffice. It is possible to inspire business models from other industries for eHealth, for example, in 2000. Parente described four e-commerce-inspired eHealth business models that were emerging at that time along with the growth of e-commerce generally [41]. E-commerce activities are probably easier to mimic from other industries than business models for health services and their complex value cocreation activities.

Not only are new business models for eHealth needed but also needed are the approaches for creating them. Admittedly, the lack of publications that discuss *how* business models can be created is not only a problem in eHealth. In general, few approaches to defining business models exist or remain cursory. Another barrier is the problem of introducing business-like thinking in health care. This continues to be a sensitive topic, as in the field of health care, the focus is the well-being of patients; thus, focusing on money is considered in a negative light because it is not patient-centered. However, with the emerging problems that health care is facing, business-like thinking could be pivotal in keeping quality health care affordable.

Future Research

We have applied and are applying the CeHRes Roadmap in several of our eHealth projects, which are all quite varied and exist in different settings ranging in complexity and size, yet all of these projects are focused on providing some form of technology that supports disease management. A few example projects that have made or are currently making use of the roadmap and, therefore, also of business modeling are shown in Textbox 9.

Textbox 9. Examples of Projects Using the CeHRes Roadmap

- Collaboration platform for cross-border infection prevention
- Setting up an antibiotic stewardship program
- Development of a teledermatology device
- Personal assistance website for diabetes care
- Prevention and quick warnings regarding the dangers of Lyme disease

XSL•FO
RenderX

All of these cases are useful for testing and improving the roadmap and are relevant to this paper. They are test cases for the current instruments for business modeling. We see that the roadmap and business modeling are applicable in all these different types of eHealth technologies, and we are working on adding instruments and evaluating current instruments. In a subsequent paper, we will give an introduction to these instruments and how they can support eHealth development. Our goal is to find robust instruments that are generic enough to be applicable for all eHealth technologies. Thus far, we have seen with our current focus groups and workshops as well as with our mapping tools that the extra effort of business modeling gives vital information not only for the implementation but also vital information with consequences for the design of the eHealth technology.

We also plan a systematic review to predetermine outcomes and effects of interventions in the antibiotic stewardship programs. After this review, we hope to assess how a literature review can be used as input for the start of the value specification by providing the outcomes and effects as general value drivers to discuss with the stakeholders.

The roadmap has been made public as a wiki (ehealthwiki.org). The goal is to provide a platform for anyone interested to collaborate on providing methods, ideas, and example cases for eHealth development as described by our roadmap. Obviously, we would also like to see contributions to the business modeling side of the roadmap.

Acknowledgments

We would like to express our thanks to Robin Kok for reviewing the draft version.

Conflicts of Interest

None declared.

References

1. Adams J, Mounib E, Pai A, Stuart N, Thomas R, Tomaszewicz P. IBM Institute for Business Value. Somers, NY: IBM Global Services; 2006. Healthcare 2015: Win-win or lose-lose? URL: http://www-935.ibm.com/services/us/gbs/bus/pdf/healthcare2015-win-win_or_lose-lose.pdf [accessed 2011-12-16] [WebCite Cache ID 63yLRYX3s]
2. Committee on Quality of Health Care in America. Institute of Medicine. Washington, DC: National Academy Press; 2001. Crossing the Quality Chasm: A New Health System for the 21st Century URL: http://www.nap.edu/openbook.php?isbn=0309072808 [accessed 2011-12-16] [WebCite Cache ID 63yLuBEp1]
3. Grossman J. Disruptive innovation in health care: Challenges for engineering. The Bridge 2008;38(1) [FREE Full text]
4. Rappa M. North Carolina State University. Raleigh, NC; 2000. Business models on the web URL: http://digitalenterprise.org/models/models.html [accessed 2011-12-16] [WebCite Cache ID 63yMlKEdP]
5. O'Reilly T. O'Reilly. Sebastopol, CA: O'Reilly Media, Inc; 2005 Sep 30. What is Web 2.0 URL: http://oreilly.com/web2/archive/what-is-web-20.html [accessed 2011-12-07] [WebCite Cache ID 63kcZNYea]
6. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
7. Ahern DK. Challenges and opportunities of eHealth research. Am J Prev Med 2007 May;32(5 Suppl):S75-S82. [doi: 10.1016/j.amepre.2007.01.016] [Medline: 17466822]
8. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
9. Jai Ganesh A. E-health-drivers, applications, challenges ahead and strategies: A conceptual framework. Indian Journal of Medical Informatics 2004;1:39-47.
10. Eng TR. eHealth research and evaluation: challenges and opportunities. J Health Commun 2002;7(4):267-272. [doi: 10.1080/10810730290001747] [Medline: 12356287]
11. Kerr C, Murray E, Stevenson F, Gore C, Nazareth I. Interactive health communication applications for chronic disease: patient and carer perspectives. J Telemed Telecare 2005;11 Suppl 1:32-34. [doi: 10.1258/1357633054461912] [Medline: 16035986]
12. LeGrow G, Metzger J. iHealth Reports. Oakland, CA: California Health Care Foundation; 2001. E-Disease Management URL: http://psg-mac43.ucsf.edu/ticr/syllabus/courses/2/2003/03/18/Lecture/readings/EDiseaseManagement.pdf [accessed 2011-12-16] [WebCite Cache ID 63yNKsoWP]
13. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]
14. Curry SJ. eHealth research and healthcare delivery beyond intervention effectiveness. Am J Prev Med 2007 May;32(5 Suppl):S127-S130. [doi: 10.1016/j.amepre.2007.01.026] [Medline: 17466817]
15. Nijland N, van Gemert-Pijnen J, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]

XSL•FO
RenderX

16. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
17. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
18. Weick K, Sutcliffe K, Obstfeld D. Organizing and the process of sensemaking. Organization Science 2005;16(4):409-421. [doi: 10.1287/orsc.1050.0133]
19. Freeman R. The stakeholder approach revisited. Zeitschrift für Wirtschafts-und Unternehmensethik 2004;5(3):228-241. [doi: 10.1287/orsc.1040.0066]
20. Carr D, Howells A, Chang M, Hirji N, English A. An integrated approach to stakeholder engagement. Healthc Q 2009;12 Spec No Ontario:62-70 [FREE Full text] [Medline: 19458512]
21. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
22. Cain M, Mittman R. iHealth Reports. Oakland, CA: California HealthCare Foundation; 2002 May. Diffusion of Innovation in Health Care URL: http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/D/PDF%20DiffusionofInnovation.pdf [accessed 2011-12-16] [WebCite Cache ID 63yNjvRpQ]
23. Agle B, Mitchell R, Sonnenfeld J. Who matters to CEOs? An investigation of stakeholder attributes and salience, corporate performance, and CEO values. Academy of Management Journal 1999;42(5):507-525.
24. Sharp H, Finkelstein A, Galal G. Stakeholder identification in the requirements engineering process. Proceedings of 10th International Workshop on Database & Expert Systems Applications 1999:387-391. [doi: 10.1109/DEXA.1999.795198]
25. Mitchell R, Agle B, Wood D. Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. Academy of Management Review 1997;22(4):853-886.
26. Prahalad CK, Ramaswamy V. Co creation experiences: The next practice in value creation. Journal of Interactive Marketing 2004;18(3):5-14.
27. Chesbrough H. Open Business Models: How to Thrive in the New Innovation Landscape. Boston, MA: Harvard Business School Press; 2006.
28. McGrath R. Business models: A discovery driven approach. Long Range Planning 2010;43:247. [doi: 10.1016/j.lrp.2009.07.005]
29. Wilson B, Athanasiou J. The Value of Healthcare IT (HIT): A Practical Approach to Discussing and Measuring the Benefits of HIT Investments. Santa Clara, CA: Intel Corporation; 2007. URL: http://www.intel.com/Assets/PDF/whitepaper/health-317997001.pdf [accessed 2011-12-16] [WebCite Cache ID 63yOSjQCw]
30. Zott C, Amit R. Business model design: An activity system perspective. Long Range Planning 2010;43:216. [doi: 10.1016/j.lrp.2009.07.004]
31. Rockart JF. The changing role of the information systems executive: A critical success factors perspective. Sloan Management Review 1982;24(1):3-13.
32. Forman E, Peniwati K. Aggregating individual judgments and priorities with the analytic hierarchy process. European Journal of Operational Research 1998;108(1):165-169.
33. Gordijn J, Akkermans H, Van Vliet H. Business modelling is not process modelling. Conceptual Modeling for E-Business and the Web 2000;1921:40-51. [doi: 10.1007/3-540-45394-6_5]
34. Osterwalder A, Pigneur Y. Business Model Generation. In: Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers (portable version). Switzerland: OSF; 2010.
35. Morris M, Schindehutte M, Allen J. The entrepreneur's business model: Toward a unified perspective. Journal of Business Research 2005;58(6):726-735. [doi: 10.1016/j.jbusres.2003.11.001]
36. Osterwalder A, Pigneur Y, Tucci C. Clarifying business models: Origins, present, and future of the concept. Communications of the Association for Information Systems 2005;16(1):1-25.
37. Magretta J. Why business models matter. Harv Bus Rev 2002 May;80(5):86-92, 133. [Medline: 12024761]
38. Doganova L, Eyquem-Renault M. What do business models do?:: Innovation devices in technology entrepreneurship. Research Policy 2009;38:1559. [doi: 10.1016/j.respol.2009.08.002]
39. Teece D. Business models, business strategy and innovation. Long Range Planning 2010;43:172. [doi: 10.1016/j.lrp.2009.07.003]
40. Shafer S, Smith H, Linder J. The power of business models. Business Horizons 2005;48(3):199-207. [doi: 10.1016/j.bushor.2004.10.014]
41. Parente ST. Beyond the hype: a taxonomy of e-health business models. Health Aff (Millwood) 2000;19(6):89-102 [FREE Full text] [Medline: 11192424]

Abbreviations

CeHRes: Center for eHealth Research and Disease Management

XSL•FO
RenderX

HCD: human-centered design
HIT: health care information technology
IT: information technology
MRSA: methicillin-resistant staphylococcus aureus

Edited by H Potts; submitted 20.10.10; peer-reviewed by B Fernando, D Silber, J Van Ooteghem; comments to author 07.01.11; revised version received 06.09.11; accepted 19.09.11; published 28.12.11.

Please cite as:
van Limburg M, van Gemert-Pijnen JEWC, Nijland N, Ossebaard HC, Hendrix RMG, Seydel ER
Why Business Modeling is Crucial in the Development of eHealth Technologies
J Med Internet Res 2011;13(4):e124
URL: http://www.jmir.org/2011/4/e124/
doi:10.2196/jmir.1674
PMID:22204896

©Maarten van Limburg, Julia EWC van Gemert-Pijnen, Nicol Nijland, Hans C Ossebaard, Ron MG Hendrix, Erwin R Seydel. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Weight Change in a Commercial Web-Based Weight Loss Program and its Association With Website Use: Cohort Study

Melinda Neve[1,3], PhD, BND; Philip J Morgan[2,3], PhD, BEd (HPE) (Hons1); Clare E Collins[1,3], PhD BSC Dip Nutr&Diet Dip Clin Epi

[1]School of Health Sciences, Faculty of Health, University of Newcastle, Callaghan, NSW, Australia

[2]School of Education, Faculty of Education and Arts, University of Newcastle, Callaghan, Australia

[3]Priority Research Centre in Physical Activity and Nutrition, University of Newcastle, Callaghan, NSW, Australia

Corresponding Author:
Melinda Neve, PhD, BND
School of Health Sciences
Faculty of Health
University of Newcastle
Priority Research Centre for Physical Activity and Nutrition, Advanced Technology Centre (Level 3)
University Drive
Callaghan, NSW, NSW
Australia
Phone: 61 61 2 49215405
Fax: 61 61 2 49212804
Email: melinda.neve@newcastle.edu.au

Abstract

Background: There is a paucity of information in the scientific literature on the effectiveness of commercial weight loss programs, including Web-based programs. The potential of Web-based weight loss programs has been acknowledged, but their ability to achieve significant weight loss has not been proven.

Objective: The objectives were to evaluate the weight change achieved within a large cohort of individuals enrolled in a commercial Web-based weight loss program for 12 or 52 weeks and to describe participants' program use in relation to weight change.

Method: Participants enrolled in an Australian commercial Web-based weight loss program from August 15, 2007, through May 31, 2008. Self-reported weekly weight records were used to determine weight change after 12- and 52-week subscriptions. The primary analysis estimated weight change using generalized linear mixed models (GLMMs) for all participants who subscribed for 12 weeks and also for those who subscribed for 52 weeks. A sensitivity analysis was conducted using the last observation carried forward (LOCF) method. Website use (ie, the number of days participants logged on, made food or exercise entries to the Web-based diary, or posted to the discussion forum) was described from program enrollment to 12 and 52 weeks, and differences in website use by percentage weight change category were tested using Kruskal-Wallis test for equality of populations.

Results: Participants (n = 9599) had a mean (standard deviation [SD]) age of 35.7 (9.5) years and were predominantly female (86% or 8279/9599) and obese (61% or 5866/9599). Results from the primary GLMM analysis including all enrollees found the mean percentage weight change was −6.2% among 12-week subscribers (n = 6943) and −6.9% among 52-week subscribers (n = 2656). Sensitivity analysis using LOCF revealed an average weight change of −3.0% and −3.5% after 12 and 52 weeks respectively. The use of all website features increased significantly ($P < .01$) as percentage weight change improved.

Conclusions: The weight loss achieved by 12- and 52-week subscribers of a commercial Web-based weight loss program is likely to be in the range of the primary and sensitivity analysis results. While this suggests that, on average, clinically important weight loss may be achieved, further research is required to evaluate the efficacy of this commercial Web-based weight loss program prospectively using objective measures. The potential association between greater website use and increased weight loss also requires further evaluation, as strategies to improve participants' use of Web-based program features may be required.

(J Med Internet Res 2011;13(4):e83) doi:10.2196/jmir.1756

XSL•FO
RenderX

KEYWORDS

Weight loss; obesity; intervention; Internet; commercial

Introduction

As the prevalence of overweight and obesity among adults continues to increase across the world [1], the need for cost-effective programs that achieve clinically important weight loss and have a broad reach are urgently needed. However, at this time there is no universally effective method of weight management that assures long-term maintenance of lost weight [2,3]. Despite this, many overweight and obese men (44%) and women (65%) report trying to lose weight [4], and many of these individuals join commercial weight loss programs [5].

The most recent systematic review of major commercial weight loss programs concluded that there was inadequate evidence to recommend their use [5], and that further randomized controlled trials (RCT) were required to provide evidence to support or refute the use of commercial weight loss programs [5]. Although controlled trials are essential to demonstrate efficacy, the results of these trials may not always be generalizable to typical enrollees of commercial programs. Therefore, studies evaluating outcomes for fee-paying commercial weight loss program participants can establish the nature of consumer engagement and the degree of weight loss that can be expected after specific periods of enrollment [6].

Commercial weight loss program providers commonly offer Web-based versions of their programs. Recent systematic reviews of Web-based weight loss interventions have highlighted the potential of these programs to achieve significant weight loss [7-10]. However, only one commercial Web-based weight loss program (ie, eDiets) has undergone rigorous testing within two RCTs, conducted in 2004 and 2007 [11,12]. The first RCT found that after 12 months, participants in the program achieved significantly less weight loss compared with participants given a self-help manual, [11]. The second RCT compared eDiets with a structured behavioral Web-based program and found that participants in the behavioral program achieved significantly greater weight loss compared with those participants randomized to eDiets after 12 months [12]. Therefore, further research is required to evaluate the weight change achieved from participation in commercial Web-based weight loss programs.

A 2010 systematic review of Web-based weight loss interventions found that greater weight loss is likely to be associated with increased use of Web-based program features [9]. This is consistent with the results of the majority of studies investigating the association between intervention exposure and outcomes, that is, greater use of program components within Web-based interventions is associated with greater weight loss or better weight loss maintenance. Program components include log-ins [12-20], self-monitoring of weight, diet and/or exercise [11,12,18,20-25], attendance at online meetings or chat sessions [12,22,25], forum posts [12,22], viewing online lessons [21], as well as overall website use [26]. Therefore, a vital component of achieving successful weight outcomes through Web-based weight loss interventions appears to be their ability to engage participants. However, we have limited knowledge of whether the association between website use and weight loss holds true for fee-paying members of commercial Web-based programs.

Therefore, the primary aim of this study was to describe the weight loss achieved by a cohort of enrollees of a commercial Web-based weight loss program among participants who subscribed to the program for 12 or 52 weeks. The secondary aim was to describe participants' use of the Web-based program overall and by percentage weight loss category and to determine if website use differed by percentage weight loss category.

Methods

Participants and Design

Participants were eligible for inclusion in the study if they paid for a subscription to the program from August 15, 2007, through May 31, 2008. To join the program, participants must have been 18 to 75 years of age and have had a body mass index (BMI) greater than or equal to 22 kg/m^2 based on self-reported height and weight. When participants enrolled, they purchased a subscription plan of 4-, 12-, 16- or 52-weeks duration. In 2007-2008, a subscription cost A$16.50 to A$79.95 per month dependant on the number of months a participant subscribed. Participants could not unsubscribe from their selected plan until the subscription timeframe had elapsed unless they had special circumstances that prevented them from completing their subscription (eg, pregnancy or financial difficulties). This study included participants who subscribed for the most popular durations of 12- or 52-weeks. Data related to free or non-consecutive subscriptions (≥ 7 days apart) were also excluded.

Characteristics of the full cohort [27] and the subgroup who subscribed for periods of 12 and 52 weeks [28] have been previously published.

The Commercial Web-Based Weight Loss Program

In 2007-2008, SP Health Co Pty Ltd (Sydney, Australia) offered a Web-based weight loss platform that was commercially available in Australia as The Biggest Loser Club (www.biggestloserclub.com.au). It was promoted as a 12-week program, but participants could choose to subscribe for longer to assist with further weight loss and/or maintenance. The self-directed program incorporated evidence-based weight management strategies and aligned with key elements of social cognitive theory [29]. Participants set a goal weight and were encouraged to work towards this target in "mini goals" (eg, 5kg or 5%). Participants were encouraged to self-monitor by reporting their weight or other body measurements via the website or short message service (SMS) and could view graphs and charts detailing their progress over time (eg, weight and waist circumference change). Participants were encouraged to weigh in once per week and received weekly reminders to do so via email or SMS during the initial 12-week program. A daily energy intake target was set based on the participant's sex, weight, height, and physical activity level to facilitate either a weight loss of 0.5 kg to 1 kg per week or maintain current

XSL•FO
RenderX

weight. Participants were encouraged to self-monitor their dietary intake and exercise using an online diary that calculated daily energy intake and expenditure. Online information in the form of weekly tutorials, fact sheets, meal, and exercise plans and weekly challenges were provided during the initial 12-week program. After 12-weeks, participants continued to receive weekly Web-based tutorials. Participants were also prompted to access the online information via a weekly email. Social support was available via a discussion board to communicate with other members.

Data Collection

All data were collected by SP Health Co, provided to the researcher in deidentified form, and included enrollment survey responses (anthropometric measures, ie, weight and height, and demographics, ie, age, gender, and postcode), subscription data (date of enrollment, date membership ceased, and subscription plans held), website use (date of log-in, online food and exercise diary entries, and posts to the discussion forum), and self-reported weight records (date of record and weight recorded). Ethics approval for the study was obtained from the University of Newcastle Human Research Ethics Committee.

Measures

Participants' characteristics were captured from the enrollment survey. Self-reported height and weight were used to calculate BMI (weight in kg/height in m^2), which was categorized as healthy, overweight, or obese using the World Health Organization's (WHO) BMI classification [30]. Reported postcodes were assigned an Index of Relative Socioeconomic Advantage and Disadvantage (IRSAD) tertile (ranked from 1 = disadvantage to 10 = advantaged) [31] as an indicator of socioeconomic status, as well as an Accessibility/Remoteness Index of Australia (ARIA) [32] to classify residential level of remoteness.

Data relating to the subscription plans participants held were used to determine whether participants enrolled for 12 or 52 weeks. The date of enrollment and the date that membership ceased were used to calculate the number of days each participant was a member of the program and, therefore, how many participants cancelled their subscription. The self-reported weight records were used to describe the number of people who weighed in each week. The self-reported weights (in kilograms) were used to determine the weight change achieved. The total number of days per week each of the website features (log-ins, food diary entries, exercise diary entries, and forum posts) were used was calculated to describe overall website use.

Data Analysis

Data analysis was undertaken using Stata 11.0 (StataCorp, College Station, Texas, USA), with *P* values less than .01 considered statistically significant. Descriptive statistics are described as means and standard deviations (SDs) for normally distributed continuous variables, medians and interquartile ranges (IQR) for nonnormal continuous data, and percentages for categorical variables.

Absolute and percentage weight change were calculated from enrollment to 12 weeks for participants who subscribed for 12 weeks and from enrollment to 52 weeks for participants who subscribed for 52 weeks. The primary analysis, to determine the weight change achieved by all program enrollees, was conducted using generalized linear mixed models (GLMMs) containing available self-reported weight records for all participants. GLMM was used because this is the preferred method for longitudinal data with missing values [33,34]. Baseline age, BMI, socioeconomic status, and remoteness were controlled for in the analyses as potential confounders.

A secondary sensitivity analysis was conducted to determine the robustness of the results from the GLMM approach. This analysis was required as GLMM are based on the assumption that missing data are missing at random, which many not be the case for data reported as part of a weight loss program. Therefore, a sensitivity analysis was conducted by imputing missing data for weight using the last observation carried forward (LOCF) method.

Spearman's rank correlations were calculated to explore associations of weight change with website use. This included the percentage weight change results from the LOCF analyses. Participants were divided into four percentage weight loss categories (weight gain, 0% to < 5% weight loss, 5% to < 10% weight loss, and ≥ 10% weight loss) based on the LOCF analysis results. The median and IQR website use was described by percentage weight loss group and differences between groups investigated using Kruskal-Wallis test for equality of populations.

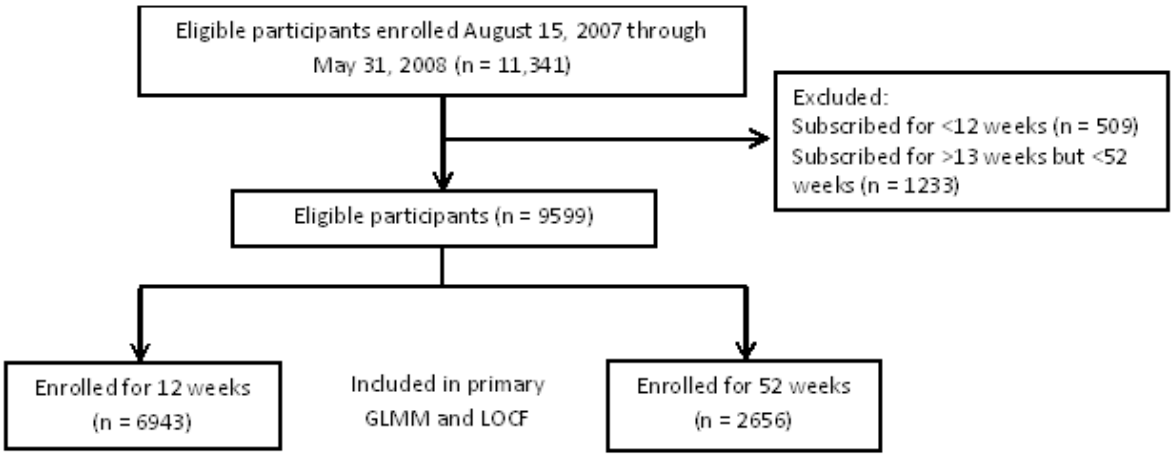
Results

Participant Characteristics

Participant flow is reported in Figure 1. A total of 11,341 participants subscribed to the program from August 15, 2007, through May 31, 2008. This study included 9599 participants; 6943 subscribed to the program for 12 weeks and 2656 subscribed for 52 weeks. Participant characteristics at enrollment have been described in detail elsewhere [27]. In summary, participants had a mean (SD) age of 35.7 (9.5) years and were predominantly female (86% or 8279/9599), obese (61% or 5866/9599), of moderate to high socioeconomic status (85% or 8022/9455 scored between 5 and 10 on IRSAD), and from major cities in Australia (75% or 7125/9456). Participants who subscribed for 12 weeks were significantly younger (35.3 years of age vs 36.7 years of age), had lower BMI (31.8 vs 35.8), were of higher socioeconomic status (39.1% vs 32.8% IRSAD 9-10), and were more likely to live in major cities of Australia (76.4% vs 72.7%), compared with 52-week subscribers. In addition, 3% (238/6943) of 12-week subscribers and 23% (605/2656) of 52-week subscribers cancelled their subscription during their subscription period due to special circumstances.

XSL•FO
RenderX

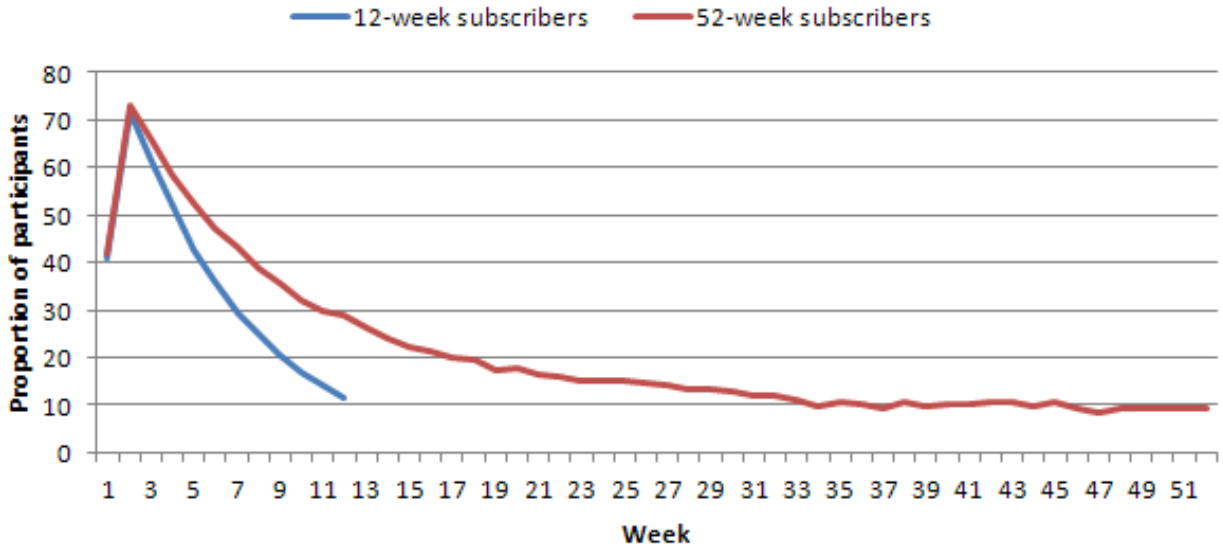
Figure 1. Participant flow through the trial.



Self-Reported Weight Record

The proportion of participants who self-reported their weight each week declined substantially over time (Figure 2). Therefore, the amount of missing weight data increased. For both 12- and 52-week subscribers, the highest proportion of participants self-reported their weight during week 2 (72% and 73%). For 12-week subscribers, only 11% (792/6943) self-reported their weight during their final week of the program (ie, 89% of participants' weight data was missing). For 52-week subscribers, the decline in the number of participants self-reporting their weight was continuous from week 2 (73%) to week 32 (12%). However, after week 32, the percentage of participants self-reporting a weight reached a plateau but remained steady at 9% to 11% until 52 weeks. Therefore, 91% (2412 /2656) of participants' weight data was missing at week 52.

Figure 2. Percent of participants who weighed in per week for 12- and 52-week subscribers.



Weight Change: Primary Analysis

Weight change results for 12- and 52-week subscribers are shown in Table 1. The GLMM gave a mean self-reported weight reduction for 12-week subscribers of −5.6 kg (95% confidence interval [CI] −5.8 kg to −5.5 kg) or −6.2% and included an average of 5.2 weekly self-reported weight records per participant. The mean self-reported weight change among 52-week subscribers was −8.4 kg (95% CI −9.0 kg to −7.8 kg) or −6.9% from the GLMM. The analysis included an average of 11.8 weekly self-reported weight records per participant.

Weight Change: Sensitivity Analysis

The sensitivity analysis using LOCF gave a mean self-reported weight loss of −2.6 kg (95% CI −2.7 kg to −2.5 kg) or −3.0%, and 21% (1479/6943) achieved greater than or equal to 5% weight loss after 12 weeks (Table 1). The sensitivity analysis using LOCF gave a mean self-reported weight change of −3.6 kg (95% CI −3.8 kg to −3.3 kg) or −3.5% from baseline to 52 weeks with 29% (777/2656) of participants achieving greater than or equal to 5% weight loss from enrollment to 52 weeks (Table 1).

XSL•FO
RenderX

Table 1. Mean (95% CI) weight change for a cohort of participants who subscribed to a commercial Web-based weight loss program for 12 or 52 weeks using GLMM and LOCF analyses

Cohort and Weight Change Measure	GLMM Analysis[a,b]	LOCF Analysis[a,b]
12-week subscribers (n = 6943)		
Absolute weight change (95% CI)	−5.6 kg (−5.8 kg to −5.5 kg)	−2.6 kg (−2.7 kg to −2.5 kg)
Percentage weight change (95% CI)	−6.2% (−6.3% to −6.1%)	−3.0% (−3.0% to −2.9%)
Percentage weight change category		
Weight gain, n (%)		423 (6.1%)
0% to < 5%, n (%)		5041 (72.6%)
5% to < 10%, n (%)		1206 (17.4%)
10% or more, n (%)		273 (3.9%)
52-week subscribers (n = 2656)		
Absolute weight change (95% CI)	−8.4 kg (−9.0 kg to −7.8 kg)	−3.6 kg (−3.8 kg to −3.3 kg)
Percentage weight change (95% CI)	−6.9% (−7.3% to −6.5%)	−3.5% (−3.8% to −3.3%)
Percentage weight change category		
Weight gain, n (%)		424 (16.0%)
0% to < 5%, n (%)		1455 (54.8%)
5% to < 10%, n (%)		475 (17.9%)
10% or more, n (%)		302 (11.4%)

[a]Difference from baseline to 12 and 52 weeks is statistically significant for all analyses ($P < .001$).

[b]Controlled for baseline age, BMI, socioeconomic status, and remoteness

Website Use

Website use for 12- and 52-week subscribers is presented in Table 2. To summarize, 12-week subscribers logged on to the website a median of 13 days. They made food entries to the Web-based diary a median of 7 days and exercise entries, a median of 3 days. The median number of days that 12-week subscribers posted to the discussion forum was zero. Among 52-week subscribers, the median number of days participants logged on was 21 days. They used the Web-based diary for food entries a median of 8 days and exercise entries a median 3 days, with a median of zero posts to the discussion forums.

Table 2. Description of 12- and 52-week subscribers' use of the website features

	12-Week Subscribers (n = 6943)		52-Week Subscribers (n = 2656)	
	Participants Who Used the Feature, n (%)	Median (IQR)	Participants Who Used the Feature, n (%)	Median (IQR)
Log-ins	6682 (96.2%)	13 (6-26)	2576 (97.0%)	21 (7-56)
Food diary entries	5244 (75.5%)	7 (1-20)	1993 (75.0%)	8 (1-34)
Exercise diary entries	4686 (67.5%)	3 (0-9)	1801 (67.8%)	3 (0-15)
Posts to the discussion forum	860 (12.4%)	0 (0-0)	1055 (39.7%)	0 (0-0)

Website Use and Weight Change

For both 12- and 52-week subscribers, percentage weight change was significantly positively correlated ($P < .001$) with the number of days each website feature was used (Table 3). The strongest correlations were found between the number of days participants logged on and weight change for 12- and 52-week subscribers. The weakest correlations were found between forum posts and weight change in both subscription groups. The strongest correlations were in the 12-week subscription group for all website features except forum posts, where the correlation between forum posts and weight change was stronger among 52-week subscribers.

XSL•FO
RenderX

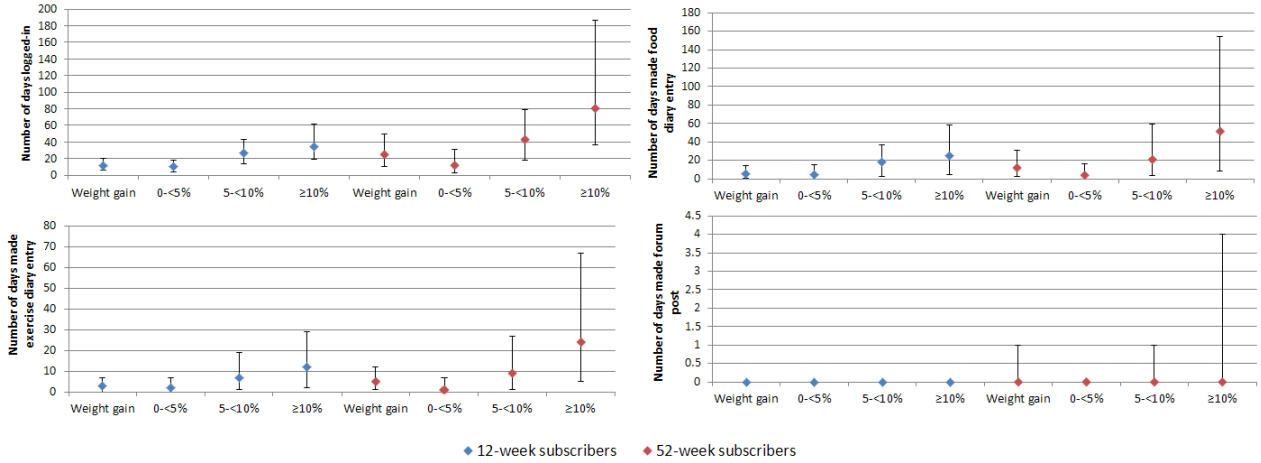
Table 3. Spearman correlations between website use and percentage weight change (kg) among 12- and 52-week subscribers

	12-Week Subscribers (n = 6943)	52-Week Subscribers (n = 2656)
	r[a]	r[a]
Log-ins	−0.55	−0.43
Food diary entries	−0.39	−0.33
Exercise diary entries	−0.38	−0.33
Forum posts	−0.12	−0.18

[a]All are statistically significant ($P < .001$)

The median number of days participants used each website feature increased significantly ($P < .001$) by category of higher percentage weight loss (Figure 3) for both 12- and 52-week subscribers. Among 12-week subscribers, those who lost 10% or more of their enrollment weight logged on a median of 34 days, made food entries to the Web-based diary 25 days, and made exercise entries 12 days, whereas those who gained weight logged on a median of 12 days, made food entries to the Web-based diary 6 days, and made exercise entries 3 days. For 52-week subscribers, those who lost 10% or more of enrollment weight logged on a median of 81 days, made food entries to the Web-based diary 52 days, and made exercise entries 24 days compared with those who gained weight, who had a median of 25 log-in days, used the Web-based food diary for food entries a median of 12 days, and made exercise entries for 5 days.

Figure 3. Median (IQR) days each website feature was used by 12- and 52-week subscribers by categories of percentage weight change.



Discussion

The primary aim of this paper was to describe the weight loss achieved by a large cohort of participants who subscribed to a commercial Web-based weight loss program for either 12 or 52 weeks. The study addresses an existing gap in the literature [5,6] by reporting weight loss outcomes in a large naturalistic cohort of commercial users of a Web-based weight loss program and its association with website feature usage. This study is one of only a small number of evaluations of commercial weight loss program cohorts and only the second to employ a robust statistical analysis as opposed to reporting results for program completers only. To the authors' knowledge, it is the first cohort study reporting outcomes from a large group of enrollees in a commercial Web-based program.

Weight Loss

Our primary analysis using GLMM indicated that both 12- and 52-week subscribers achieved statistically significant weight loss. Mean weight loss also exceeded the benchmark (≥ 5%) for clinically important weight loss and improvement in weight-related morbidity, particularly incidence of type 2 diabetes mellitus [35,36]. Furthermore, 21% of 12-week subscribers and 29% of 52-week subscribers achieved a weight loss greater than or equal to 5%, based on the results from the LOCF analysis.

However, the sensitivity analysis at both time points demonstrated less weight loss compared with the GLMM. GLMM assumes that any data missing from the model follow the same trajectory as the included data (in this case weekly weight change). As the average number of weekly weight records included was low and most people self-reported their weekly weight within the initial weeks of the program only, the GLMM results may be biased toward those who self-reported more weekly weights. It is likely that the participants who did not enter their weights were the less successful participants. This is supported by our previous findings that participants with poor eating and activity habits were more likely to stop using the program [28]. Furthermore, it is also likely that the rate of weight loss during the initial weeks of the program was higher compared with the later stages of the program; therefore, the trajectory of the GLMM may also be biased toward higher self-reported weight loss. Therefore, the true weight loss achieved by all participants at each time point is likely to be

XSL•FO
RenderX

somewhere in the range between the GLMM and LOCF results (ie, −3.0% to −6.2% at 12 weeks and −3.5% to −6.9% at 52 weeks). Therefore, further research is required to confirm or refute these findings prospectively and objectively in a clinical research trial.

Results from the only two RCTs conducted using another commercial Web-based weight loss program, eDiets, reported a mean percentage weight change of −2.8% [12] and −1.1% [11] after 12 months. Both eDiets and the commercial Web-based program evaluated in this study included many of the components that have been suggested as key elements of successful Web-based weight management programs [37,38], such as self-monitoring, feedback, and social support. However, eDiets also included additional features not available in the program evaluated in the current study, such as online meetings, peer-mentoring [11,12], and face-to-face sessions with a psychologist [11]. It was expected that these additional program components would lead to greater weight change. However, the mean weight change achieved in the current study was greater. This is potentially due to the increased capabilities of the Internet since the first study was conducted and/or differences in study design. So, although both programs provided similar features, those in the current study may potentially have been more engaging, easier, and/or faster for participants to use, reducing the burden to adhere.

Website Use and Weight Loss

The second aim of the paper was to describe participants' use of the Web-based program and its features and to determine if website use was associated with degree of weight loss.

The study demonstrated a significant positive correlation between the number of times each website feature was used and weight change. Therefore, the results support previous research [9] suggesting that ongoing engagement with Web-based weight loss programs may enhance weight loss in the long-term. Given this association, strategies are required to encourage participants to use Web-based weight loss programs consistently to ensure that the majority of participants are given the opportunity to achieve clinically important weight loss.

However, at the group level, the average use of the commercial Web-based weight management program features appears to be low and inconsistent. The majority of subscribers log on and try the Web-based diary at least once; however, engagement decreases quite fast. This is demonstrated by the initial decline in weekly self-reported weight records over time for both 12- and 52-week subscribers and is consistent with other public health interventions delivered via the Internet, where usage declines after the initial weeks of the intervention [39].

As this commercial Web-based weight loss program is self-directed, the intensity or frequency of website use is not prescribed. Therefore, this study provides valuable data and insight into what level of website use may be feasible and, more importantly, what level is required to be effective in achieving weight change in a commercial setting. Interestingly, participants who achieved significant weight loss did not use the website unrealistically or excessively. For example, those who achieved greater than or equal to 10% weight loss from baseline to 12 weeks logged on approximately 40% of the possible days (34 days out of 84) and used the Web-based diary 30% of possible days (25 days out of 84). These findings suggest that developing program targets for weekly or monthly website use and for specific program features may increase usage and enhance weight loss, thus facilitating achievement of participants' weight loss goals. However, to identify optimal exposure to the website overall, as well as individual website features, further investigation of the differences in use at different stages of the program and its association with weight loss is required. For example, this study demonstrates that participants who achieved greater than or equal to 10% weight loss from baseline to 12 weeks logged on approximately 40% of the possible days (34 days out of 84), whereas those who achieved the same percentage weight change from baseline to 52 weeks logged on approximately 22% of the days (81/365). Therefore, further research is needed to investigate the relationship between patterns of website use over time and the weight loss achieved at different time points.

Limitations

There are several important considerations when interpreting the weight change results. First, the weight change results are based on self-report, and weight is commonly underreported [40]. However, self-reported weight recorded by participants of a Web-based weight loss program has been found to be accurate compared with measured weight [41]. Second, a notable number of weekly weight records were missing, as the weight data was entered voluntarily by participants as part of their program participation and many participants failed to do this. To address this, statistical analyses were conducted using GLMM. GLMMs are among the most robust statistical methods available as these models are less influenced by the bias introduced because of missing data. Additionally, a large number of individual weekly weight records were included in each analysis (31,228 and 36,339) allowing the analyses to be strongly powered. Therefore, the results from the statistical analysis provide us with an indication of the weight loss achieved by a cohort of enrollees of a commercial Web-based weight loss program. However, due to the low level of website use and, therefore, the very small number of participants still self-reporting their weight at the end of their subscription period, further research is required to confirm or refute these findings and to identify ways to increase participant engagement with the program

The website use data and the reported associations with weight change also have some limitations to be noted. First, the study did not consider use of all website features as these data were not available at the time of the study. Additional data concerning the use of all features (eg, weekly tutorials), as well as more detailed data on the reported features (eg, whether participants read the forum posts) would help to better understand participants' engagement with the website and the relationship between weight loss and website use. Second, the analysis to determine if greater website use was associated with enhanced weight loss relied on the results of the LOCF analysis. As previously stated, the true weight loss achieved by all participants is likely to be somewhere in the range between the GLMM and LOCF results. Third, although an association

XSL•FO
RenderX

between website use and weight loss was demonstrated, a large number of other factors may have influenced participants' website use and/or weight loss (eg, self motivation, intention to change, and other weight loss strategies) that were not evaluated in this study. Therefore, the association between website use and weight loss must also be confirmed prospectively in an objective manner.

Conclusion

In summary, this research provides important data on an underevaluated weight loss program medium in a large number of commercial program users. The weight loss achieved by 12- and 52-week subscribers of a commercial Web-based weight loss program is likely to be in the range of the primary and sensitivity analysis results. This suggests that, on average, clinically important weight loss may be achieved. The findings support the need for further research to evaluate the efficacy of Web-based weight loss programs and to assist in the development of strategies to increase participants' ongoing use of Web-based program features.

Acknowledgments

We acknowledge the work of Ben Noblet in retrieving the data, Anna Crook and Penelope Jones from SP Health for assistance with managing the data set, and Patrick McElduff for assistance with design and implementation of the statistical analyses. MJ Neve was funded by an Australian Postgraduate Award scholarship and a scholarship top-up from SP Health Co Pty Ltd. CE Collins is supported by an Australian National Health and Medical Research Council Career Development Award research fellowship.

Conflicts of Interest

MJ Neve received a postgraduate scholarship top-up from SP Health Co. CE Collins has been a consultant dietitian to SP Health Co. PJ Morgan and CE Collins hold an Australian Research Council (ARC) Linkage project grant that is evaluating a weight loss program with SP Health Co.

References

1. Ogden CL, Yanovski SZ, Carroll MD, Flegal KM. The epidemiology of obesity. Gastroenterology 2007 May;132(6):2087-2102. [doi: 10.1053/j.gastro.2007.03.052] [Medline: 17498505]
2. Wadden TA, Crerand CE, Brock J. Behavioral treatment of obesity. Psychiatr Clin North Am 2005 Mar;28(1):151-70, ix. [doi: 10.1016/j.psc.2004.09.008] [Medline: 15733617]
3. Levy RL, Finch EA, Crowell MD, Talley NJ, Jeffery RW. Behavioral intervention for the treatment of obesity: strategies and effectiveness data. Am J Gastroenterol 2007 Oct;102(10):2314-2321. [doi: 10.1111/j.1572-0241.2007.01342.x] [Medline: 17561967]
4. Andreyeva T, Long MW, Henderson KE, Grode GM. Trying to lose weight: diet strategies among Americans with overweight or obesity in 1996 and 2003. J Am Diet Assoc 2010 Apr;110(4):535-542. [doi: 10.1016/j.jada.2009.12.029] [Medline: 20338279]
5. Tsai AG, Wadden TA. Systematic review: an evaluation of major commercial weight loss programs in the United States. Ann Intern Med 2005 Jan 4;142(1):56-66. [Medline: 15630109]
6. Womble LG, Wang SS, Wadden TA. Commercial and self-help weight loss programs. In: Wadden TA, Stunkard AJ, editors. Handbook of Obesity Treatment. New York, NY: The Guilford Press; 2002.
7. Weinstein PK. A review of weight loss programs delivered via the Internet. J Cardiovasc Nurs 2006;21(4):251-8; quiz 259. [Medline: 16823276]
8. Saperstein SL, Atkinson NL, Gold RS. The impact of Internet use for weight loss. Obes Rev 2007 Sep;8(5):459-465. [doi: 10.1111/j.1467-789X.2007.00374.x] [Medline: 17716303]
9. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
10. Arem H, Irwin M. A review of web-based weight loss interventions in adults. Obes Rev 2011 May;12(5):e236-e243. [doi: 10.1111/j.1467-789X.2010.00787.x] [Medline: 20804523]
11. Womble LG, Wadden TA, McGuckin BG, Sargent SL, Rothman RA, Krauthamer-Ewing ES. A randomized controlled trial of a commercial internet weight loss program. Obes Res 2004 Jun;12(6):1011-1018. [doi: 10.1038/oby.2004.124] [Medline: 15229342]
12. Gold BC, Burke S, Pintauro S, Buzzell P, Harvey-Berino J. Weight loss on the web: A pilot study comparing a structured behavioral intervention to a commercial program. Obesity (Silver Spring) 2007 Jan;15(1):155-164 [FREE Full text] [doi: 10.1038/oby.2007.520] [Medline: 17228043]
13. Bennett GG, Herring SJ, Puleo E, Stein EK, Emmons KM, Gillman MW. Web-based weight loss in primary care: a randomized controlled trial. Obesity (Silver Spring) 2010 Feb;18(2):308-313 [FREE Full text] [doi: 10.1038/oby.2009.242] [Medline: 19696764]

XSL•FO
RenderX

14. Cussler EC, Teixeira PJ, Going SB, Houtkooper LB, Metcalfe LL, Blew RM, et al. Maintenance of weight loss in overweight middle-aged women through the Internet. Obesity (Silver Spring) 2008 May;16(5):1052-1060 [FREE Full text] [doi: 10.1038/oby.2008.19] [Medline: 18309301]
15. McConnon A, Kirk SF, Cockroft JE, Harvey EL, Greenwood DC, Thomas JD, et al. The Internet for weight control in an obese sample: results of a randomised controlled trial. BMC Health Serv Res 2007;7:206 [FREE Full text] [doi: 10.1186/1472-6963-7-206] [Medline: 18093289]
16. Petersen R, Sill S, Lu C, Young J, Edington DW. Effectiveness of employee internet-based weight management program. J Occup Environ Med 2008 Feb;50(2):163-171. [doi: 10.1097/JOM.0b013e31815c6cf6] [Medline: 18301173]
17. 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] [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]
18. 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 [FREE Full text] [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
19. 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]
20. Jonasson J, Linné Y, Neovius M, Rössner S. An Internet-based weight loss programme -- a feasibility study with preliminary results from 4209 completers. Scand J Public Health 2009 Jan;37(1):75-82. [doi: 10.1177/1403494807085066] [Medline: 18684784]
21. Wing RR, Crane MM, Thomas JG, Kumar R, Weinberg B. Improving weight loss outcomes of community interventions by incorporating behavioral strategies. Am J Public Health 2010 Dec;100(12):2513-2519. [doi: 10.2105/AJPH.2009.183616] [Medline: 20966375]
22. Webber KH, Tate DF, Michael Bowling J. A randomized comparison of two motivationally enhanced Internet behavioral weight loss programs. Behav Res Ther 2008 Sep;46(9):1090-1095. [doi: 10.1016/j.brat.2008.06.008] [Medline: 18675402]
23. Polzien KM, Jakicic JM, Tate DF, Otto AD. The efficacy of a technology-based system in a short-term behavioral weight loss intervention. Obesity (Silver Spring) 2007 Apr;15(4):825-830 [FREE Full text] [doi: 10.1038/oby.2007.584] [Medline: 17426316]
24. Morgan PJ, Lubans DR, Collins CE, Warren JM, Callister R. The SHED-IT randomized controlled trial: evaluation of an Internet-based weight-loss program for men. Obesity (Silver Spring) 2009 Nov;17(11):2025-2032 [FREE Full text] [doi: 10.1038/oby.2009.85] [Medline: 19343018]
25. 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. [doi: 10.1038/oby.2004.40] [Medline: 14981225]
26. Funk KL, Stevens VJ, Appel LJ, Bauck A, Brantley PJ, Champagne CM, et al. Associations of internet website use with weight change in a long-term weight loss maintenance program. J Med Internet Res 2010;12(3):e29 [FREE Full text] [doi: 10.2196/jmir.1504] [Medline: 20663751]
27. Neve MJ, Morgan PJ, Collins CE. Participant characteristics and reach of a commercial web-based weight loss program. Nutrition & Dietetics 2010;67(4):267-274. [doi: 10.1111/j.1747-0080.2010.01474.x]
28. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. J Med Internet Res 2010;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]
29. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.
30. World Health Organization. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser 2000;894:i-xii, 1. [Medline: 11234459]
31. Australian Bureau of Statistics. Information Paper: An Introduction to Socio-Economic Indexes for Areas (SEIFA). 2006. URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/2039.0/ [accessed 2011-09-21] [WebCite Cache ID 62Nk9hkLt]
32. Australian Bureau of Statistics. Australian Standard Geographical Classification (ASGC). Canberra, Australia: Australian Bureau of Statistics; 2008.
33. Chakraborty H, Gu H. A mixed model approach for intent-to-treat analysis in longitudinal clinical trials with missing values. Research Triangle Park, NC: RTI Press; 2009 Mar. URL: http://www.rti.org/pubs/mr-0009-0904-chakraborty.pdf [accessed 2011-09-21] [WebCite Cache ID 61rAwC4PX]
34. Mallinckrodt CH, Watkin JG, Molenberghs G, Carroll RJ. Choice of the primary analysis in longitudinal clinical trials. Pharmaceutical Statistics 2004;3(3):161-169. [doi: 10.1002/pst.124]
35. National Health and Medical Research Council. Clinical Practice Guidelines for the Management of Overweight and Obesity in Adults. Canberra, Australia: National Health and Medical Research Council; 2003. URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/7AF116AFD4E2EE3DCA256F190003B91D/$File/adults.pdf [accessed 2011-09-25] [WebCite Cache ID 61xjmOI8d]

XSL•FO
RenderX

36. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 7;346(6):393-403 [FREE Full text] [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
37. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938. [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]
38. Bensley RJ, Brusk JJ, Rivas J. Key principles in internet-based weight management systems. Am J Health Behav 2010;34(2):206-213. [Medline: 19814600]
39. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
40. Gorber SC, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. Obes Rev 2007 Jul;8(4):307-326. [doi: 10.1111/j.1467-789X.2007.00347.x] [Medline: 17578381]
41. Harvey-Berino J, Krukowski RA, Buzzell P, Ogden D, Skelly J, West DS. The accuracy of weight reported in a web-based obesity treatment program. Telemed J E Health 2011 Sep 1. [doi: 10.1089/tmj.2011.0032] [Medline: 21882997]

Abbreviations

ARIA: Accessibility/Remoteness Index of Australia
BMI: body mass index
CI: confidence interval
GLMM: generalized linear mixed model
ISRAD: Index of Relative Socioeconomic Advantage and Disadvantage
IQR: interquartile range
LOCF: last observation carried forward

Edited by G Eysenbach; submitted 14.02.11; peer-reviewed by K Funk, S Robroek; comments to author 25.04.11; revised version received 20.06.11; accepted 24.06.11; published 12.10.11.

Please cite as:
Neve M, Morgan PJ, Collins CE
Weight Change in a Commercial Web-Based Weight Loss Program and its Association With Website Use: Cohort Study
J Med Internet Res 2011;13(4):e83
URL: http://www.jmir.org/2011/4/e83/
doi:10.2196/jmir.1756
PMID:21993231

©Melinda Neve, Philip J Morgan, Clare E Collins. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 12.10.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of Tailoring in an Internet-Based Intervention for Smoking Cessation: Randomized Controlled Trial

Silje C Wangberg[1,2], CPsych, PhD; Olav Nilsen[3], MSc; Konstantinos Antypas[3], MSc; Inger Torhild Gram[3,4], MD, PhD

[1]Regional Centre on Substance Abuse, University Hospital of North Norway, Narvik, Norway

[2]Department of Clinical Medicine, University of Tromsø, Tromsø, Norway

[3]Norwegian Centre of Integrated Care and Telemedicine, University Hospital of North Norway, Tromsø, Norway

[4]Department of Community Medicine, University of Tromsø, Tromsø, Norway

Corresponding Author:
Silje C Wangberg, CPsych, PhD
Regional Centre on Substance Abuse
University Hospital of North Norway
PO Box 385
Narvik, NO-8505
Norway
Phone: 47 76 96 65 00
Fax: 47 76 96 68 79
Email: siljecw@gmail.com

Abstract

Background: Studies suggest that tailored materials are superior to nontailored materials in supporting health behavioral change. Several trials on tailored Internet-based interventions for smoking cessation have shown good effects. There have, however, been few attempts to isolate the effect of the tailoring component of an Internet-based intervention for smoking cessation and to compare it with the effectiveness of the other components.

Objective: The study aim was to isolate the effect of tailored emails in an Internet-based intervention for smoking cessation by comparing two versions of the intervention, with and without tailored content.

Methods: We conducted a two-arm, randomized controlled trial of the open and free Norwegian 12-month follow-up, fully automated Internet-based intervention for smoking cessation, slutta.no. We collected information online on demographics, smoking, self-efficacy, use of the website, and participant evaluation at enrollment and subsequently at 1, 3, and 12 months. Altogether, 2298 self-selected participants aged 16 years or older registered at the website between August 15, 2006 and December 7, 2007 and were randomly assigned to either a multicomponent, nontailored Internet-based intervention for smoking cessation (control) or a version of the same Internet-based intervention with tailored content delivered on the website and via email.

Results: Of the randomly assigned participants, 116 (of 419, response rate = 27.7%) in the intervention group and 128 (of 428, response rate = 29.9%) in the control group had participated over the 12 months and responded at the end of follow-up. The 7-day intention-to-treat abstinence rate at 1 month was 15.2% (149/982) among those receiving the tailored intervention, compared with 9.4% (94/999) among those who received the nontailored intervention ($P < .001$). The corresponding figures at 3 months were 13.5% (122/902) and 9.4% (84/896, $P = .006$) and at 12 months were 11.2% (47/419) and 11.7% (50/428, $P = .91$). Likewise, the intervention group had higher self-efficacy and perceived tailoring at 1 and 3 months. Self-efficacy was found to partially mediate the effect of the intervention.

Conclusion: Tailoring an Internet-based intervention for smoking cessation seems to increase the success rates in the short term, but not in the long term.

(J Med Internet Res 2011;13(4):e121) doi:10.2196/jmir.1605

KEYWORDS

Smoking cessation; tailored feedback; email

XSL•FO
RenderX

Introduction

The Internet seems to provide a promising setting for combining the ability to reach a lot of smokers with good effectiveness with a low cost per smoker. In a recent Cochrane review of 20 randomized and quasi-randomized trials on Internet-based interventions for smoking cessation, Civljak and colleagues [1] concluded that some Internet-based interventions can assist smoking cessation. Interventions appropriately tailored to the users and with frequent automated contact seemed most promising, although the results were inconsistent. Shahab and McEwen [2] concluded in their meta-analysis of Internet-based interventions for smoking cessation that the tailored interventions increased 6-month abstinence rates by 17% (95% confidence interval [CI], 12%–21%) compared with the nontailored intervention.

A tailored intervention is one that is adapted to the characteristics of the individual, and it is typically based on responses to a questionnaire. The main ways of tailoring can be classified into personalization, adaptation, and feedback [3]. Personalization refers to making references to the recipient in the text such as first name, age, gender, or hometown. Adaptation concerns the content of the text itself, which can be tailored according to a variety of theories. Health psychological models often form the core of adaptive tailored interventions. Self-efficacy is one of the theoretical constructs that have shown the most consistent effects as a result of tailoring [4]. The third method of tailoring, feedback, is a widely used feature of tailoring in which the recipient is informed about scores on a scale and how the score can be interpreted. In recent, more complex tailoring, these features are often combined, and the components of the Internet-based intervention may also be tailored.

Although the literature suggests that tailoring is an important part of Internet-based interventions for supporting health behavioral change, we do not know how important it is compared with other components, such as discussion forums, personal quitting plans, and diaries, or how these components might interact. One way of studying these relationships is to compare a full intervention with a version where one of the components, such as tailoring, has been removed. Strecher et al [5] compared a tailored Internet-based intervention for smoking cessation with a nontailored Internet-based intervention and found that after 12 weeks, continuous abstinence rates (using the number of users who had logged on at least once as the denominator) were 22.8% in the tailored group compared with 18.1% (odds ratio = 1.34) in the nontailored group. Etter [6] compared two versions of the smoking-cessation program Stop-Tabac.ch, where the control group received an online report tailored to a number of variables, whereas the intervention group received a similar report that was somewhat targeted to a reasonable stage of change according to their smoking status, but otherwise fixed in terms of the tailoring variables (eg, self-efficacy was set as low for all and attitude toward smoking was set as positive for all). The result was a report that might have actually been tailored for some by chance, but not for all the participants. At the immediate follow-up 48 hours later, it was found that 12% (intention-to-treat [ITT]) of the smokers in both groups had made a 24-hour quit attempt. Plausible explanations of this lack of increased effect in the tailored group include an unclear control condition with targeting to stage of change and the potential for both actual and pseudotailoring, in addition to the very short follow-up time period.

We aimed to isolate the effect of tailored feedback in a multicomponent Internet-based intervention for smoking cessation through randomly allocating participants to one of two versions of our Internet-based intervention: one with tailored feedback, or one that was otherwise similar but without the tailored feedback. The purpose of the study was to examine, in a 12-month randomized controlled trial, whether the 7-day abstinence rates would differ between those receiving the tailored intervention and those who did not. We also wanted to explore whether tailoring would result in improved self-efficacy and more use of the website.

Methods

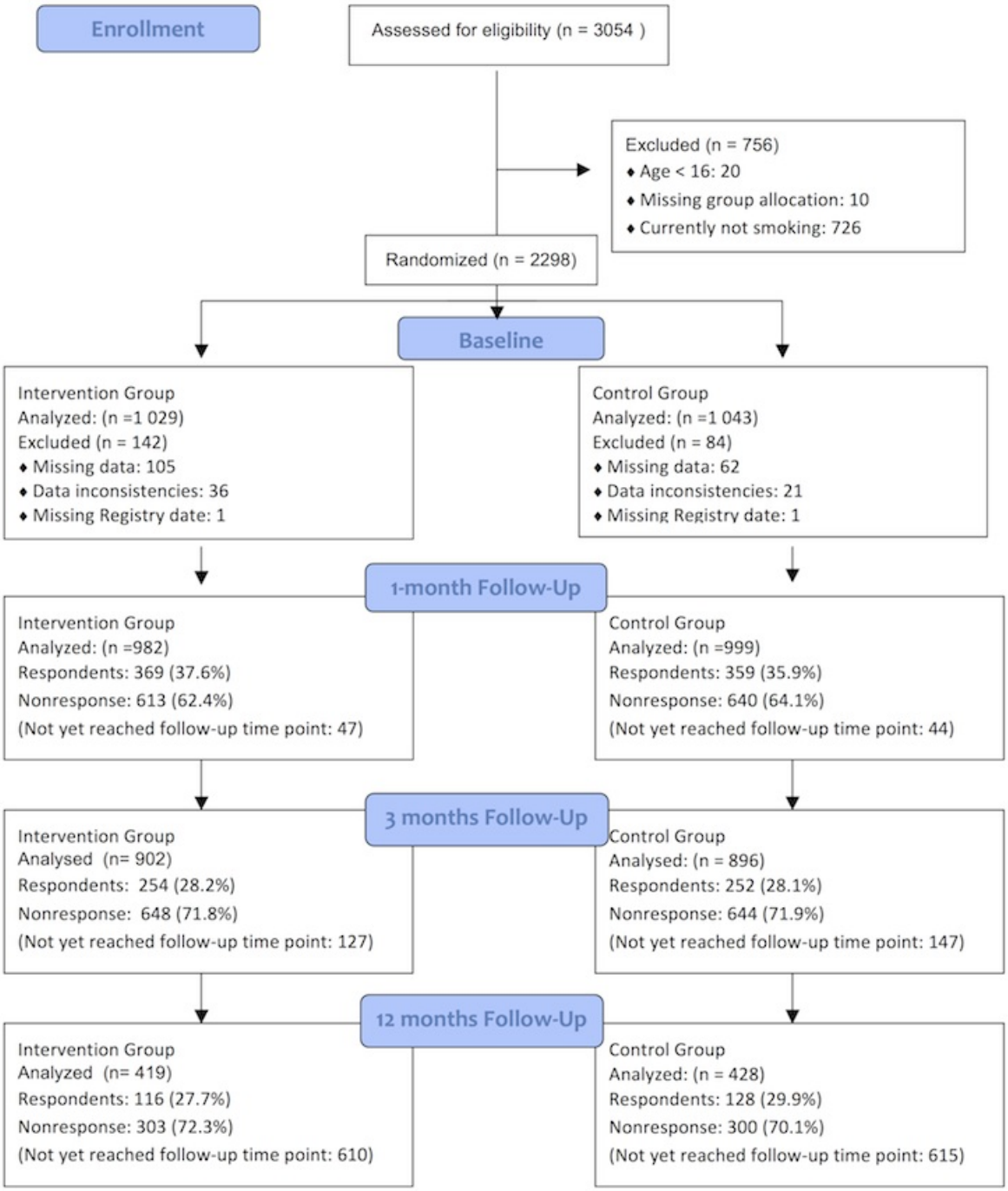
Study Design and Participants

The study was a two-arm, 12-month, randomized controlled, Internet-based trial with continuous recruitment and data collection. The allocation ratio was 1:1. The intervention arm in the trial received tailored messages in addition to the basic functionality of the Internet-based intervention for smoking cessation, while the control arm did not. Enrollment started on August 15, 2006 and ended December 7, 2007. The study was approved by the Regional Ethics Committee for North Norway (REK-NORD) and the Norwegian Privacy Ombudsman for Research. The randomized controlled trial was initiated before trial registration became customary in Norway, and therefore does not have a trial identification number.

The Intervention website was announced as a new and free service to aid in smoking cessation in the local and national media. All participants agreeing to the informed consent form were subsequently automatically allocated through use of an online random number generator to the intervention or control arm (for the informed consent form, see Multimedia Appendix 1). Altogether 3054 visitors registered to use the Norwegian website slutta.no. The front page displayed the logos of the Norwegian Directorate of Health's Quitline, the Norwegian Cancer Society, and the Norwegian Centre for Telemedicine (for a screenshot of the front page of the intervention, see Multimedia Appendix 2). Registration required providing a unique email address, so potentially, using several email addresses, a person could have registered more than once. Among registrants, 30 were excluded (20 because they were under age 16 years and 10 because of missing group allocation). Another 726 registrants had already quit smoking and were excluded from the current analyses. Among the 2298 participants who smoked at enrollment, 1029 were randomly assigned to the intervention and 1043 to the control arm (Figure 1).

XSL•FO
RenderX

Figure 1. Participant flowchart.



The participants filled in an extensive questionnaire at the time of enrollment. This included information on such demographic variables as gender, age, education, and work situation. The participants also provided a quit date and an email address. Further, they completed a smoking-cessation maintenance self-efficacy questionnaire, reported on smoking behavior, and stated their motivation for cessation. The tailored messages were created on the basis of these questionnaires and were sent to the intervention group on their personal webpage and by email (for a screenshot of My Page, see Multimedia Appendix

XSL•FO
RenderX

2). Participants in the control group did not get any messages on their webpage and only emails containing notifications and reminders for the follow-up questionnaires.

Both arms received an email message with a link to a questionnaire on self-efficacy and smoking behavior at 1, 3, and 12 months after the date of enrollment. On completion of the questionnaire, the participants could enter a draw to win prizes such as books and T-shirts. Nonresponders received up to three reminders.

Internet-Based Intervention for Smoking Cessation (slutta.no): Basic Functionality

The intervention consisted of multiple intervention components and was intended for long-term follow-up. The website included static information on the dangers of smoking, general advice on smoking cessation, and information about the website. In addition there were interactive tests for nicotine addiction, type of smoker (stress smoker, comfort smoker, etc), and motivation level.

There was an emphasis on creating opportunities for social interaction using a discussion forum, a guestbook, and a personal diary (Multimedia Appendix 3). The participants could invite friends and other participants to support them by leaving messages in the guestbook, and to publish their own smoking-cessation diary. There were also some community features: participants could click on other participants' nicknames in the forum and thereby get a specific profile with some information about the other participant, for example. The possibilities to interact were only as described above, as there were no opportunities for synchronous communication through chat or private messaging between the participants.

Tailoring

During the 12-month intervention, the participants in the intervention group received up to 150 tailored messages. The first message was sent 14 days before, and the last, 12 months after, the quit date. The frequency of the messages varied over the course of the 12 months, as they were sent out on a defined number of days before or after the quit date. In the beginning messages were sent daily, then the frequency was decreasing slowly during the first 3 months with a substantial drop-off 3 months after the quit date.

The tailoring was set up on the basis of several different types of variables. Personalization-, adaption-, and feedback-type tailoring were all used to varying degrees. Table 1 lists examples of variables for tailoring. The self-efficacy messages were more specifically about confidence in refraining from relapsing in different situations, also known as maintenance self-efficacy [7]. In concordance with several stage and process models of health behavioral change, such as the Health Action Process Approach [7], we aimed at providing these as preparation to transition from conscious behavioral change (action) to lifestyle integration (maintenance). In this intervention we did not assess where participants were in their process through a questionnaire, but we did send maintenance self-efficacy messages to those with a low maintenance self-efficacy at 3 months past their quit date. There was no other tailoring based on a health psychological stage or process approach in the current intervention. Besides the messages concerning addiction, the rest concerning benefits of quitting smoking, social support, etc, were evenly distributed over the year, with decreasing frequency.

The tailored messages could also be retrieved from a calendar on the participant's My Page. Other tailoring features on this page included a personalized greeting, feedback on number of smoke-free days and the amount of money saved, and a list of the reasons the participant had entered for wanting to quit smoking.

Table 1. Examples of the tailoring the participants in the intervention group received

Variable	Question	Answer example	Message example (sent relative to quit date)
Personalization	What would you like us to call you?	Jane	+365 days: Congratulations, Jane! Today you have been smoke-free for a year!
Quit date	When do you intend to quit?	[Date]	+5 days: There is no longer nicotine present in your body.
Step-down	Would you like to do a step-down of your smoking?	Yes, through smoke-free zones	–10 days: Create a smoke-free room in your home.
Self-efficacy	How confident do you feel about refraining from smoking when angry or upset?	1 = Not confident at all	Immediately on screen: Try to calm down instead of smoking when angry or upset. Relaxation techniques are one effective way to do it, and can be done quickly and discreetly, wherever and whenever, once you have practiced them. Another method is distraction. You can take a walk, read the paper, or play a game.
Main occupation	Are you currently working?	Yes, Working full time	–2 days: Consider which situations at work tempt you to smoke.
Social pressure	Do your friends smoke?	Yes, all of them	+58 days: Watch out! Some might like it if you fail. It could make them feel better.
Motivation	What is your most important reason for quitting?	I want to save money	+71 days: Try to calculate how much money you have saved. It might make you proud!
Social support	Would you like to tell others that you are quitting smoking?	Yes	–13 days: Tell your friends and family that you plan to quit.

XSL•FO
RenderX

Measures

Data were gathered on age, gender, education, perceived tailoring, perceived usefulness of the website, self-efficacy, and smoking. Education was rated on a 5-point scale: 1, ≤9 years of total education; 2, 10–11 years; 3, 12 years; 4, 13–16 years; and 5, ≥17 years. Motivation was assessed with a single question, "How strong is your motivation for quitting smoking?" The participant answered on a 4-point scale ranging from "very weak" to "very strong." Previous research has found that a similar single-item measure for motivation had comparable predictive validity to a multi-item instrument [8].

Data on the use of the interventions were gathered through Web logging. The number of log-ins and time spent at the site (in minutes) per user were registered. At the 1-month follow-up, the participants were asked whether they would recommend the site to a friend and to rate from a list of intervention components the one that they found the most useful.

Smoking behavior was assessed at the baseline and at 1-, 3-, and 12-month follow-ups as 7-day abstinence rates through the question "Have you during the last 7 days had a smoke, even just a single puff?"

Data on smoking-cessation maintenance self-efficacy were gathered at baseline and after use of the site for 1 month and 3 months, using the 12-item Smoking Self-Efficacy Questionnaire [9,10]. The 12 items consist of 6 items addressing perceived ability to refrain from smoking in the face of internal stimuli such as when nervous, and another 6 items addressing external stimuli such as when with other smokers. The scale has previously been shown to predict smoking cessation, to be negatively related to number of cigarettes smoked, and to have the ability to discriminate between stages of change [10]. The same study [10] found that the internal consistency was alpha = .94 for the internal subscale and alpha = .89 for the external. At baseline the items were rated on a 5-point scale that was later recoded into a 6-point scale (2 = 2.25, 3 = 3.5, 4 = 4.75, 5 = 6, ELSE = Copy), and on a 6-point scale at 1- and 3-month follow-up (1 = not confident at all, to 6 = completely confident). This recoding was necessary because in the tailoring questionnaires, which provided the baseline data, we used the original 5-point response format of the SEQ-12 [9] while, for purposes of conformity of response format in the evaluation questionnaires, we used a 6-point scale here.

Perceived tailoring was assessed with 4 items from Dijkstra [3] evaluating to what extent the user feels that the information is adapted to his or her personal situation. Agreement with these 4 items was rated on a 6-point scale ranging from 1, completely disagree, to 6, totally agree.

Statistical Analyses

We based our a priori sample size estimation on a paper by Strecher [11], pointing out that previous computer-based smoking-cessation intervention studies have found group differences in abstinence rates of about 2%. Using abstinence rates at 12 months that only slightly exceeded no intervention (6% and 8%) and a 1-sided test without continuity correction at a .05 alpha level and with 90% power, we needed a total sample of 2787. Also, we expected to have to raise the number of participants recruited further by 40%–60%, that is, to around 4000, because of the high dropout often observed in Internet-based interventions [12].

No items had more than 5% missing data at the baseline; we therefore assumed missing data to be missing completely at random. On the variables *self-efficacy* and *perceived tailoring*, we replaced the missing data with values imputed by the expectation maximum likelihood algorithm in SPSS version 16.0 (IBM Corporation, Somers, NY, USA) before analysis of variance. Internal consistency was measured by Cronbach alpha. Nonresponse on 7-day abstinence was dealt with by counting all participants with missing data as smokers (ITT). We compared the ITT quit rates with the quit rates for responders only.

Differences in dichotomous baseline characteristics and in abstinence rates between groups at all time points were analyzed with a regular chi-square test. Group differences in continuous variables were analyzed with *t* test. The Mann-Whitney *U* test was used for comparing the usage of the intervention between groups, as these distributions were nonnormal. Effect sizes for group differences at the different time points were calculated as relative risk.

Mediation was tested using an approach developed by Preacher and Hayes [13] using their SPSS macro [14]. Bootstrapping (5000 samples) with bias correction and acceleration was used to create a 95% CI around the point estimate of the indirect effect, with an interval not including zero indicating a significant indirect effect.

Results

Participant Characteristics and Response Rates

Both recruitment and data collection were continuous, and were maintained right until the point of data extraction. This implies that at the time of data extraction, some users of the intervention had been followed up for a few days and others for the full 12 months. In Figure 1 those participants not having had the possibility to reach the next follow-up time point are indicated as "Not yet reached follow-up time point," while the true nonresponders are indicated by "Nonresponse". The overall response rate was 36.8% (728/1981) after 1 month, 28.1% (506/1798) after 3 months, and 28.8% (244/847) after 12 months. There were no significant differences in response rates between the two groups at any time point (1 month: $\chi^2_1 = 0.58$, $P = .45$; 3 months: $\chi^2_1 < 0.001$, $P = .99$; 12 months: $\chi^2_1 = 0.51$, $P = .48$).

Overall among the participants, 72.26% (1497/2072) were female, mean age was 37 years, 17.1% (353/2072) had 17 or more years of education, mean motivation score was 3 (range 1–4), mean self-efficacy score was 34 (range 0–60), mean number of cigarettes smoked per day was 16.

As can be seen in Table 2, there were no significant differences between the intervention and the control group at baseline. Furthermore, no group differences in demographics were found at the follow-up time points.

XSL•FO
RenderX

Table 2. Baseline comparisons

	Intervention group (n = 1029)	Control group (n = 1043)	Test statistic	*P* value
Female			$\chi^2_1 = 1.37$	.24
n (%)	732 (71.1%)	766 (73.4%)		
95% CI[a]	68.3%–73.8%	70.8%–76.1%		
Age (years)			$t_{2013} = 0.94$	.35
Mean	37.3	36.9		
95% CI	36.7–38.0	36.2–37.5		
Range	16–71	16–68		
Education (years)			$\chi^2_4 = 3.22$	.52
≤9, n (%)	51 (5%)	54 (5.2%)		
95% CI	3.6%–6.4%	3.2%–6.5%		
10–11, n (%)	157 (15.3%)	188 (18%)		
95% CI	13.2%–17.5%	15.5%–20.5%		
12, n (%)	188 (18.3%)	190 (18.2%)		
95% CI	15.9%–20.7%	15.9%–20.5%		
13–16, n (%)	455 (44.2%)	436 (41.8%)		
95% CI	41.3%–47.2%	39%–44.7%		
≥17, n (%)	178 (17.3%)	175 (16.8%)		
95% CI	14.9%–19.8%	14.5%–19%		
Occupational status[b]			$\chi^2_5 = 1.78$	.88
Full-time employment, n (%)	610 (59.3%)	634 (60.8%)		
95% CI	56.4%–62.2%	57.8%–63.7%		
Part-time employment, n (%)	135 (13.1%)	130 (12.5%)		
95% CI	11.2%–15.4%	10.5%–14.4%		
Unemployed, n (%)	36 (3.5%)	28 (2.7%)		
95% CI	2.4%–4.7%	1.7%–3.7%		
Student, n (%)	149 (14.5%)	148 (14.2%)		
95% CI	12.3%–16.6%	12.1%–16.1%		
Retired, n (%)	27 (2.6%)	25 (2.4%)		
95% CI	1.7%–3.7%	1.5%–3.5%		
Cigarettes per day			$t_{2064} = –0.29$	.77
Mean	16.1	16.2		
95% CI	15.6–16.5	15.7–16.6		
Living with someone			$\chi^2_1 = 4.22$	.04
n (%)	797 (77.5%)	846 (81.1%)		
95% CI	74.9%–80.1%	78.8%–83.6%		
Motivation score			$t_{2069} = 0.87$	.38
Mean	2.96	2.93		
95% CI	2.91–3.00	2.89–2.97		
Self-efficacy score			$t_{2064} = 1.36$	.17
Mean	32.6	32		

XSL•FO
RenderX

	Intervention group (n = 1029)	Control group (n = 1043)	Test statistic	*P* value
95% CI	32–33.2	31.4–32.6		

[a] Confidence interval.

[b] There was also an ”Other” category not shown in the table.

Use of the Intervention

Table 3 displays the time spent on different activities at the website according to study arm. The intervention group had logged on more times *(P* = .03) and had used the site more overall *(P* = .02). In more detail, the intervention group had used My Page more *(P* = .03) than the control group had. The most used component of the intervention was the discussion forum, followed by My Page, while the Facts section was used much less. More detailed analyses on the use of the intervention over time can be found in Wangberg et al [15].

Table 3. Number of log-ins and minutes of use overall for some of the core components of the intervention by group

	Group	Median	IQR[a]	*Z* score	*P* value
Number of log-ins overall	Intervention (n = 1029)	3	5		
	Control (n = 1043)	2	4	4.54	<.001
Minutes spent at site overall	Intervention	93	159		
	Control	68	107	5.46	<.001
Minutes spent in discussion forum	Intervention	6	27.5		
	Control	6	29	0.92	.36
Minutes spent at My Page	Intervention	7	13		
	Control	6	9	2.21	.027
Minutes spent reading Facts	Intervention	0	1		
	Control	0	1	3.33	.001

[a] Interquartile range is a measure of variation for the median, which equals the difference between the third and the first quartile.

User Evaluation

In the intervention group, 88.4% (320/362, 95% CI, 84.7–91.3) of the users stated that they would recommend the site to a friend, compared with 71.8% (255/355, 95% CI, 66.9–76.3, *P* < .001) in the control group. Further, in the intervention group, 34.0% (123/362, 95% CI, 29.3–39.0) of the users ranked the tailored emails as the most useful intervention component, compared with 6% (21/355, 95% CI, 3.9–8.9, *P* < .001) in the control group (who did not receive any emails besides one with username and password upon registration and emails with links to follow-up questionnaires). In the intervention group, 10% (37/362, 95% CI, 7.5–13.8) of the users ranked general information as the most useful component, compared with 22% (79/355, 95% CI, 18.2–26.9, *P* < .001) in the control group, while 15% (55/362, 95% CI, 11.9–19.3) of the users in the intervention group ranked the discussion forum as the most useful component, compared with 21% (73/355, 95% CI, 16.7–25.1, *P* = .06) in the control group. The remaining nominations were evenly spread over the 10 other functions the user could choose as the most useful.

Manipulation Check: Perceived Tailoring

The *perceived tailoring* scale was found to have good internal consistency at 1-month (alpha = .92) and 3-month (alpha = .94) follow-up. Table 4 shows mean scores on perceived tailoring by group at 1- and 3-month follow-ups. The intervention group had higher perceived tailoring scores at both time points (*P*s < .001).

XSL•FO
RenderX

Table 4. Perceived tailoring scores by group at follow-up

Time point	Intervention group	Control group	*t* test	*P* value
1 month			$t_{715} = 4.50$	<.001
Mean	15.91	14.22		
95% CI[a]	15.45–16.40	13.68–14.79		
n	369	359		
3 months			$t_{502} = 4.59$	<.001
Mean	15.45	13.37		
95% CI[a]	14.85–16.09	12.72–14.04		
n	254	252		

[a] Confidence interval.

Smoking Cessation

Table 5 shows that the ITT 7-day abstinence rate at 1 month was 15.2% (149/982) among those receiving the tailored intervention, compared with 9% (94/999) among those who did not ($P < .001$). The corresponding figures at 3 months were 13.5% (122/902) and 9% (84/896, $P = .006$) and at 12 months were 11% (47/419) and 12% (50/428, $P = .91$). The same group differences were found looking at responders only (Table 5).

Table 5. Group 7-day abstinence rates

Analytic strategy	Time point	Intervention group		Control group				
		Percentage (n/total)	95% CI[a]	Percentage (n/total)	95% CI[a]	χ^2_1	*P* value	RR[b] (95% CI[a])
All nonresponders counted as smokers (intention-to-treat)	1 month	15.2% (149/982)	13.1–17.6	9% (94/999)	7.8–11.4	15.3	<.001	1.61 (1.27–2.06)
	3 months	13.5% (122/902)	11.5–15.9	9% (84/896)	7.6–11.5	7.6	.006	1.44 (1.11–1.87)
	12 months	11% (47/419)	8.5–14.6	12% (50/428)	9.0–15.1	0.05	.91	0.96 (0.66–1.40)
Responders only	1 month	40.4% (149/369)	35.4–45.4	26% (94/359)	21.6–30.8	16.5	<.001	1.54 (1.25–1.91)
	3 months	48.0% (122/254)	41.9–54.2	33% (84/252)	27.5–39.2	11.3	.001	1.44 (1.16–1.79)
	12 months	41% (47/116)	31.5–49.6	39% (50/128)	30.5–47.6	0.1	.82	1.03 (0.76–1.41)

[a] Confidence interval.

[b] Relative risk.

Secondary Outcome: Self-efficacy

Both the internal (alpha = .93) and the external (alpha = .86) self-efficacy subscales were found to have good internal consistency at 1-month follow-up. Table 6 shows the mean scores for self-efficacy at all follow-up time points. Self-efficacy was higher for the intervention group at 1- ($P = .01$) and 3-month ($P = .002$) follow-ups, but not after 1 year ($P = .58$), paralleling the results for the main outcome.

XSL•FO
RenderX

Table 6. Self-efficacy score by group at follow-up

Time point	Intervention group	Control group	*t* test	*P* value
1 month			$t_{626} = 3.60$	.01
Mean	41.57	38.36		
95% CI[a]	40.42–42.71	37.03–39.70		
n	369	359		
3 months			$t_{336} = 3.15$	.002
Mean	42.45	38.69		
95% CI[a]	40.82–43.98	36.98–40.38		
n	254	252		
12 months			$t_{211} = 0.51$	.58
Mean	39.59	38.60		
95% CI[a]	36.96–42.24	35.79–41.57		
n	116	128		

[a] Confidence interval.

Test of Mediation: Self-efficacy and Perceived Tailoring

We performed a mediational analysis (n = 386) with group as the independent variable, 7-day abstinence at 3 months as the dependent variable, and self-efficacy and perceived tailoring at 1 month as the proposed mediators. The total effect of group on abstinence rates at 3-month follow-up was 0.66 (Wald $\chi^2_1 = 9.82$, $P = .002$). Self-efficacy accounted for an indirect effect of 0.33 (95% CI, 0.08–0.60), while perceived tailoring did not have a significant indirect effect (point estimate = 0.003, 95% CI, –0.09 to 0.10). The remaining direct effect of group on abstinence rates was 0.58 (Wald $\chi^2_1 = 5.82$, $P = .02$).

Discussion

The results show that both 7-day abstinence rates and self-efficacy for smoking cessation were higher among those in the tailored intervention group at 1- and 3-month follow-ups, but not at the 12-month follow-up. The short-term results are consistent with previous studies [2].

We found that the intervention group had used the intervention more. One of the ways that tailoring may lead to higher smoking-cessation success is through providing a higher dose of the intervention. A previous study has shown that tailored emails increased adherence to the same Internet-based smoking-cessation intervention, but only up until 5 months [15]. Simple dose–response relationships have been found previously [16-18], but Danaher and colleagues [19] did not find a mediational effect of program exposure when controlling for self-efficacy, suggesting that the issue is not as simple as mere quantity of exposure to the intervention. Like Danaher and colleagues [19], we also found that self-efficacy partially mediated the effect of the intervention.

More participants in the intervention group than in the control group would recommend the intervention to a friend, with the tailored emails being ranked as the most useful feature of the intervention. In comparison, the participants in the control group (who did not receive the tailored emails) found the generic information and the discussion forum to be the most useful features.

The intervention group, which was the only one receiving tailored content, reported higher scores on perceived tailoring at 1- and 3-month follow-ups. We did not find that perceived tailoring mediated the effect of the intervention. Perceived program relevance (and amount of the materials read) have previously been found to mediate the effect of a tailored Internet-based smoking-cessation program [20], and an experimental study has even shown that perceived tailoring can account for the effect in a placebo tailored condition [21]. This was further supported by a later study where Webb and colleagues [22] were able to increase the effect of tailoring further by using expectancy (that tailored content is superior to generic) priming.

The main strengths of the current Internet-based smoking-cessation trial were a high sample size and repeated measurement. As this was an effectiveness trial, the results have higher external validity. Only age, access to the Internet, and willingness to set a quit date during the next 3 months were inclusion criteria for the present study, thus providing more relevant information for implementation in a real-world setting outside of strongly controlled clinical trials. At the same time, however, the representativeness of the study was compromised by the fact that the sample was self-selected. The results, thus, cannot be generalized to all people pursuing smoking cessation. Especially, the findings are less generalizable to men, since women in this study, as in previous ones [2], were overrepresented. Women generally tend to use the Internet more for health purposes than men do, and possible reasons for this include women's traditional caretaking role and greater preferences for social support [23]. Our sample also had a relatively high educational attainment, and we are currently

XSL•FO
RenderX

running a trial (clinical trial #NCT011030427) on whether the use of short message service (SMS) can increase use of the intervention by those with lower educational attainment. A study we did on delivering diabetes information via SMS suggested that the short format and push delivery might increase attention to and comprehension of the information [24].

A limitation of our study was that we were not able to separate receiving tailored content from receiving emails per se. Another limitation that this study shares with many other Internet-based interventions [14,25] is a high attrition and, thus, low response rate at follow-up assessments. It is likely that some of the participants through interactions in the discussion forum noted that they had not received the "full" version—for example, did not receive any advice by email. Despite this, we did not find differential attrition, and reached similar conclusions concerning the main outcome whether we used the ITT strategy of counting all nonresponders as smoking or analyzed just the responders. A follow-up study of nonresponders to a quitline study indicates that the ITT yields too low actual quit rates, as many of the people they followed up were (still) abstinent [26].

Furthermore, as seen from the records of website use in this study, and in Internet-based interventions generally, the consistency of delivery is often high, although the amount of time spent with the intervention can vary greatly between participants, with some of them barely visiting the site at all, as also seen in previous research [25,27-29].

Conclusions

This randomized controlled trial found that tailoring an Internet-based intervention for smoking cessation increases success rates in the short term, but not in the long term.

Acknowledgments

This study was supported in part by EXTRA funds from the Norwegian Foundation for Health and Rehabilitation and by funds from the Norwegian Directorate of Health. The intervention presented in the current paper was developed in collaboration between the Norwegian Directorate of Health's Quitline, the Norwegian Cancer Society, and the Norwegian Centre for Telemedicine, and is currently owned and run by the Norwegian Directorate of Health. The authors especially express thanks to Inger Merete Skarpaas at the Norwegian Cancer Society for her contributions to developing the intervention. Some of the tailoring questions came from the CD-ROM Dr SmokeFree developed by Dr Hans Gilljam and owned by Preventive Media Sweden AB. The second author was funded by a doctoral fellowship from the Northern Norway Regional Health Authority, grant no. TFP743-08.

Conflicts of Interest

The first author headed the development of the intervention tested in this paper.

Authors' Contributions

All authors took part in interpreting data and revising content. SCW additionally designed and carried out the study, led the analysis and interpretation of data, and drafted the manuscript. OL and KA additionally analysed and interpreted data, and drafted parts of the manuscript.

Multimedia Appendix 1

Informed consent form for the intervention Opptur/www.slutta.no. (Translated from Norwegian).

[JPG File, 923KB - jmir_v13i4e121_app1.jpg]

Multimedia Appendix 2

Screenshot of the front page of the intervention Opptur/www.slutta.no.

[JPG File, 97KB - jmir_v13i4e121_app2.jpg]

Multimedia Appendix 3

Screenshot of "My Page" where, among other functionality such as a diary and a guest book, tailored feedback was displayed on the intervention Opptur/www.slutta.no.

[JPG File, 147KB - jmir_v13i4e121_app3.jpg]

References

1. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010(9):CD007078. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
2. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. Addiction 2009 Nov;104(11):1792-1804. [doi: 10.1111/j.1360-0443.2009.02710.x] [Medline: 19832783]

XSL•FO
RenderX

3. Dijkstra A. Working mechanisms of computer-tailored health education: evidence from smoking cessation. Health Educ Res 2005 Oct;20(5):527-539 [FREE Full text] [doi: 10.1093/her/cyh014] [Medline: 15701665]
4. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
5. Strecher VJ, Shiffman S, West R. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. Addiction 2005 May;100(5):682-688. [doi: 10.1111/j.1360-0443.2005.01093.x] [Medline: 15847626]
6. Etter JF. Comparing computer-tailored, internet-based smoking cessation counseling reports with generic, untailored reports: a randomized trial. J Health Commun 2009;14(7):646-657. [doi: 10.1080/10810730903204254] [Medline: 19851916]
7. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
8. Sciamanna CN, Hoch JS, Duke GC, Fogle MN, Ford DE. Comparison of five measures of motivation to quit smoking among a sample of hospitalized smokers. J Gen Intern Med 2000 Jan;15(1):16-23. [Medline: 10632829]
9. Etter JF, Bergman MM, Humair JP, Perneger TV. Development and validation of a scale measuring self-efficacy of current and former smokers. Addiction 2000 Jun;95(6):901-913. [Medline: 10946439]
10. Christie DH, Etter JF. Validation of English-language versions of three scales measuring attitudes towards smoking, smoking-related self-efficacy and the use of smoking cessation strategies. Addict Behav 2005 Jun;30(5):981-988. [doi: 10.1016/j.addbeh.2004.10.002] [Medline: 15893093]
11. Strecher VJ. Computer-tailored smoking cessation materials: a review and discussion. Patient Educ Couns 1999 Feb;36(2):107-117. [Medline: 10223016]
12. Cobb NK, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. Nicotine Tob Res 2005 Apr;7(2):207-216. [doi: 10.1080/14622200500055319] [Medline: 16036277]
13. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008 Aug;40(3):879-891. [Medline: 18697684]
14. Hayes AF. afhayes.com. My Macros and Code for SPSS and SAS (SPSS Macro for Bootstrapping Specific Indirect Effects in Multiple Mediation Models) URL: http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html [accessed 2011-11-11] [WebCite Cache ID 637MsJoeR]
15. Wangberg SC, Bergmo TS, Johnsen JA. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
16. Graham AL, Cobb NK, Raymond L, Sill S, Young J. Effectiveness of an internet-based worksite smoking cessation intervention at 12 months. J Occup Environ Med 2007 Aug;49(8):821-828. [doi: 10.1097/JOM.0b013e3180d09e6f] [Medline: 17693778]
17. Japuntich SJ, Zehner ME, Smith SS, Jorenby DE, Valdez JA, Fiore MC, et al. Smoking cessation via the internet: a randomized clinical trial of an internet intervention as adjuvant treatment in a smoking cessation intervention. Nicotine Tob Res 2006 Dec;8 Suppl 1:S59-S67. [Medline: 17491172]
18. An LC, Perry CL, Lein EB, Klatt C, Farley DM, Bliss RL, et al. Strategies for increasing adherence to an online smoking cessation intervention for college students. Nicotine Tob Res 2006 Dec;8 Suppl 1:S7-12. [Medline: 17491165]
19. Danaher BG, Smolkowski K, Seeley JR, Severson HH. Mediators of a successful web-based smokeless tobacco cessation program. Addiction 2008 Oct;103(10):1706-1712. [doi: 10.1111/j.1360-0443.2008.02295.x] [Medline: 18715238]
20. Strecher VJ, Shiffman S, West R. Moderators and mediators of a web-based computer-tailored smoking cessation program among nicotine patch users. Nicotine Tob Res 2006 Dec;8 Suppl 1:S95-101. [Medline: 17491176]
21. Webb MS, Simmons VN, Brandon TH. Tailored interventions for motivating smoking cessation: using placebo tailoring to examine the influence of expectancies and personalization. Health Psychol 2005 Mar;24(2):179-188. [doi: 10.1037/0278-6133.24.2.179] [Medline: 15755232]
22. Webb MS, Hendricks PS, Brandon TH. Expectancy priming of smoking cessation messages enhances the placebo effect of tailored interventions. Health Psychol 2007 Sep;26(5):598-609. [doi: 10.1037/0278-6133.26.5.598] [Medline: 17845111]
23. Wangberg SC. Digital outdistancing: socio-economic status differences in use of the Internet: what could it mean for health disparities? In: Evans CM, editor. Internet Issues: Blogging, the Digital Divide and Digital Libraries. Hauppauge, NY: Nova Science Publishers; 2010.
24. Wangberg SC, Arsand E, Andersson N. Diabetes education via mobile text messaging. J Telemed Telecare 2006;12 Suppl 1:55-56. [doi: 10.1258/135763306777978515] [Medline: 16884582]
25. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
26. 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;5:52 [FREE Full text] [doi: 10.1186/1471-2458-5-52] [Medline: 15910682]
27. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008;2:97-105 [FREE Full text] [Medline: 19920949]
28. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]

XSL•FO
RenderX

29. Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11 [FREE Full text] [doi: 10.2196/jmir.9.2.e11] [Medline: 17513282]

Abbreviations

CI: confidence interval
ITT: intention-to-treat
SMS: short message service

Edited by G Eysenbach; submitted 06.07.10; peer-reviewed by JF Etter, J , J Seeley; comments to author 15.12.10; revised version received 13.09.11; accepted 25.09.11; published 15.12.11.

Please cite as:
Wangberg SC, Nilsen O, Antypas K, Gram IT
Effect of Tailoring in an Internet-Based Intervention for Smoking Cessation: Randomized Controlled Trial
J Med Internet Res 2011;13(4):e121
URL: http://www.jmir.org/2011/4/e121/
doi:10.2196/jmir.1605
PMID:22169631

©Silje C Wangberg, Olav Nilsen, Konstantinos Antypas, Inger Torhild Gram. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of an Interactive, Web-Delivered System to Increase Provider–Patient Engagement in Smoking Cessation

Rajani S Sadasivam[1,2], PhD; Kathryn Delaughter[1,2], MA; Katie Crenshaw[3], JD, MSEd; Heather J Sobko[4], BSN, PhD; Jessica H Williams[5], MPH; Heather L Coley[6], MPH; Midge N Ray[7], MSN; Daniel E Ford[8], MD, MPH; Jeroan J Allison[1], MD, MS; Thomas K Houston[1,2,9], MD, MPH

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

[2]VA eHealth Quality Enhancement Research Initiative, Bedford VAMC, Bedford, MA, United States

[3]Division of Continuing Medical Education, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[4]Division of Gerontology, Geriatrics and Palliative Care, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[5]Division of Clinical Immunology and Rheumatology, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[6]Division of General Internal Medicine, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[7]Department of Health Services Administration, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States

[8]Division of General Internal Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States

[9]Center for Health Quality, Outcomes & Economic Research (CHQOER), Bedford VAMC, Bedford, MA, United States

Corresponding Author:
Rajani S Sadasivam, PhD
Division of Health Informatics and Implementation Science
Quantitative Health Sciences
University of Massachusetts Medical School
55 Lake Ave North
Worcester, MA, 01545
United States
Phone: 1 508 856 8923
Fax: 1 508 856 8993
Email: rajani.sadasivam@umassmed.edu

Abstract

Background: Patient self-management interventions for smoking cessation are effective but underused. Health care providers do not routinely refer smokers to these interventions.

Objective: The objective of our study was to uncover barriers and facilitators to the use of an e-referral system that will be evaluated in a community-based randomized trial. The e-referral system will allow providers to refer smokers to an online smoking intervention during routine clinical care.

Methods: We devised a four-step development and pilot testing process: (1) system conceptualization using Delphi to identify key functionalities that would overcome barriers in provider referrals for smoking cessation, (2) Web system programming using agile software development and best programming practices with usability refinement using think-aloud testing, (3) implementation planning using the nominal group technique for the effective integration of the system into the workflow of practices, and (4) pilot testing to identify practice recruitment and system-use barriers in real-world settings.

Results: Our Delphi process (step 1) conceptualized three key e-referral functions: (1) Refer Your Smokers, allowing providers to e-refer patients at the point of care by entering their emails directly into the system, (2) practice reports, providing feedback regarding referrals and impact of smoking-cessation counseling, and (3) secure messaging, facilitating provider–patient communication. Usability testing (step 2) suggested the system was easy to use, but implementation planning (step 3) suggested several important approaches to encourage use (eg, proactive email cues to encourage practices to participate). Pilot testing (step 4) in 5 practices had limited success, with only 2 patients referred; we uncovered important recruitment and system-use barriers (eg, lack of study champion, training, and motivation, registration difficulties, and forgetting to refer).

XSL•FO
RenderX

Conclusions: Implementing a system to be used in a clinical setting is complex, as several issues can affect system use. In our ongoing large randomized trial, preliminary analysis with the first 50 practices using the system for 3 months demonstrated that our rigorous preimplementation evaluation helped us successfully identify and overcome these barriers before the main trial.

Trial: Clinicaltrials.gov NCT00797628; http://clinicaltrials.gov/ct2/show/NCT00797628 (Archived by WebCite at http://www.webcitation.org/61feCfjCy)

(J Med Internet Res 2011;13(4):e87) doi:10.2196/jmir.1721

KEYWORDS

Smoking cessation; general practice; family practice; public health informatics; user interfaces; randomized controlled trial; health services research; web-based services

Introduction

Smoking is the number 1 behavioral health problem and preventable cause of death in the United States [1-5]. Among its innumerable morbidities, smoking is responsible for approximately one-third of all cancer deaths [6]. Patient self-management interventions that can easily be disseminated, such as self-help materials, computer-tailored printouts, interactive voice-response systems, quitlines, and, more recently, smoking-cessation websites [4,7-14] can potentially engage much greater numbers of smokers [15]. Unfortunately, these interventions are underused [16]. For example, as few as 3.5% of adult smokers access quitlines per year [17]. These patient self-management interventions are often deployed as public health interventions and are not well connected to clinical medicine.

Because the majority of smokers (70%) see a provider at least once per year [18], point-of-care referrals could greatly increase use of publicly available self-management smoking-cessation interventions. A recent study using proactive fax referrals to quitlines demonstrated an increased number of patients using these services [19]. Although clinical providers report limited time and competing demands as barriers to referring patients to smoking-cessation resources, they also acknowledge the role of a single source of referral, additional support, referral coordinators, and reimbursement for tobacco counseling in aiding the intervention process [20]. A system seamlessly linking the physicians, nurses, and patients within a clinical microsystem may be more effective in reducing barriers to physician referrals. Further, increasing standard protocols, data collection, and feedback between individuals in the microsystem can maximize patient-centered care [21-23].

This paper describes the preimplementation evaluation of a provider e-referral system (ReferASmoker.org). ReferASmoker.org will be used in a nationwide randomized trial that will recruit 160 primary care physician practices and test the e-referral functions [24]. A system intended to be used in a clinical setting must overcome the barriers that may impede its success. These barriers may be software usability issues or problems integrating with the standard processes of care. Our "how-to" report demonstrates how small, rigorously conducted, multistep preimplementation evaluation can positively affect the success of the larger study. Our preliminary analysis in the main trial shows that our evaluation approach successfully identified many barriers in the study's formative stages, and we were able to overcome them before the main study trial.

Methods

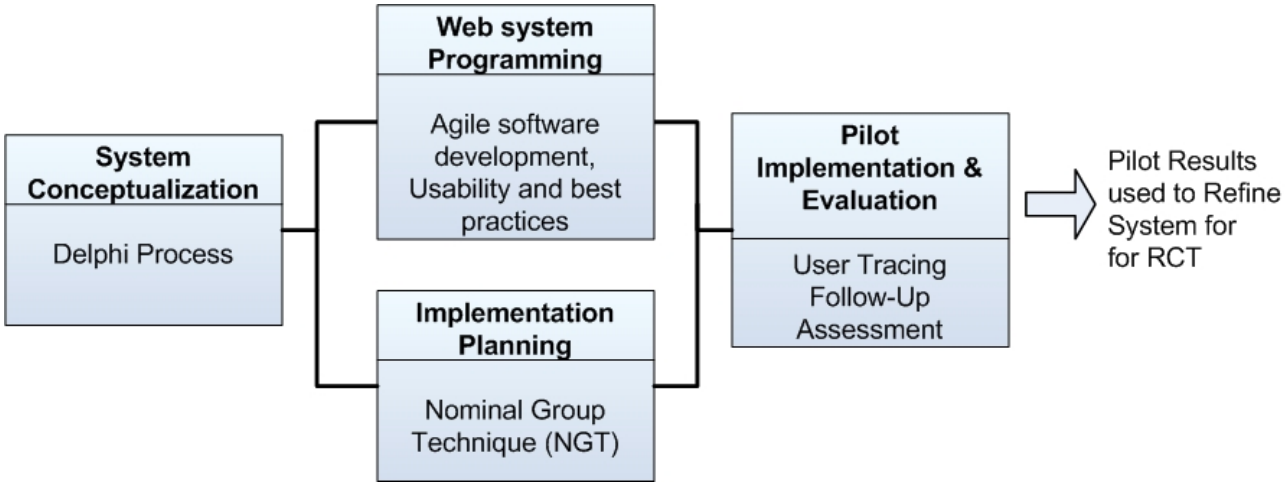
ReferASmoker.org is a point-of-care e-referral portal that allows providers to e-refer smoking patients to an online smoking-cessation portal. The ReferASmoker.org system (http://www.ReferASmoker.org) can be accessed using the email address reviewer@nih.grant and the password "review".

Study Design

Our four-step usability and pilot testing approach consisted of (1) system conceptualization using the Delphi technique to identify key functionalities that would overcome barriers in provider referrals for smoking cessation, (2) Web system programming and refinement using agile methodology and think-aloud usability testing, (3) implementation planning using the nominal group technique (NGT) for the effective deployment of the system in practices, and (4) pilot testing to identify practice recruitment and system-use barriers (Figure 1).

XSL•FO
RenderX

Figure 1. Development stages of ReferASmoker.org (RCT = randomized controlled trial).



Setting and Sample

For system conceptualization, we recruited experts (clinical, informatics, and tobacco control) from multiple academic institutions. For usability and pilot testing, we selected practices that would represent the sample in our planned randomized trial. Thus, physicians and nurses from community-based practices across several states in the United States were recruited. For our implementation planning sessions, we recruited physicians from a university setting. Our study was approved by the institutional review boards at the University of Alabama at Birmingham, the Johns Hopkins University School of Medicine, and the University of Massachusetts Medical School.

Phase 1: System Conceptualization

To conceptualize the system, we used a modified Delphi process [25,26], a systematic forecasting method for reaching consensus regarding prediction of usability and feasibility. It is a useful communication strategy that provides a structured process for the reliable and creative exploration of ideas suitable for decision making. Controlled opinion feedback sessions are used to establish expert consensus without certain social interactive behaviors that can hinder opinion forming in a typical group discussion [25].

A panel of 8 experts that included physicians and psychologists with expertise in health services, tobacco control, and informatics participated in the Delphi process. Our goals were to identify the major limitations of current smoking-cessation systems, along with identifying areas to refine in order to maximize physician engagement in the referral of patients to smoking-cessation resources within our system. We conducted three face-to-face discussions over a period of 3 weeks, and in-between email discussions augmented the process. One investigator (TKH) was responsible for synthesizing a literature review and presenting to the panel in the first face-to-face meeting. The same investigator was responsible for summarizing meeting minutes, distributing them by email, and then organizing the email discussions for the next round of the face-to-face discussions in the Delphi process.

Phase 2: Programming and Usability testing

Agile Software Development

Agile software development was used to iteratively strategize and plan the programming of the ReferASmoker.org e-referral system. Unlike the traditional approach of specifying system requirements fully at the outset of development and then undertaking programming, the system is developed in units after an overall strategy is formulated. In each agile phase, a short-term goal is set for developing a unit of the system, followed by team development of the unit, including requirements, design, programming, and testing. Agile software development is advantageous because developers can adapt to changing requirements based on the short-term goal setting and collaboration. This approach has also been demonstrated to reduce development time and risk [27].

Web System Programming

The ReferASmoker.org Web-based system was programmed using Microsoft's ASP.NET version 3.5 (Microsoft Corporation, Redmond, WA, USA) and C# technology. Microsoft SQL Server version 2000 was used as the database. We used programming best practices in the form of design patterns and modular architecture. Design patterns have been used over the years to solve software development problems. Originally introduced by the Gang of Four [28], these design patterns have evolved, and many are being used in developing Web systems.

Frameworks make it easier to use patterns. Specifically, we used the Web Client Software Factory (WCSF) version February 2008 [29], which is a .Net-based framework introduced by Microsoft. In the WCSF, the Web user interface is programmed using the model-view-presenter (MVP) design pattern [30]. The MVP pattern splits the Web interface into three layers: (1) a *model* that defines the data to be displayed or acted upon in the user interface, (2) a *view* that displays the model and routes user commands (events) to the presenter, and (3) a *presenter* that acts upon the model and the view such as formatting the data for display in the view. The modular approach of MVP makes it easier to modify the Web layer without affecting other areas of the system and to unit test the system for programming errors. In addition to the use of MVP in the Web layer, WCSF divides the rest of the system into business modules and foundational

XSL•FO
RenderX

modules. Business modules guide the programming of the business logic of the system. The foundational modules are used to program the data access and reusable functions of the system. The modular approach of WCSF makes it easier to make programming changes to the system, as each layer is only loosely connected to the others. This approach also makes it easier to independently test each layer for programming errors using mock data.

To implement data access, we used the combination of NHibernate and Castle ActiveRecord frameworks (version Release Candidate 1) [31,32]. These frameworks guide consistent and structured data access from the database using object-relational mapping (ORM). ORM is a technique that maps the relational data structure of the database into an object-oriented structure [33]. Castle ActiveRecord leverages NHibernate functions and implements the active-record pattern [34,35], a database-related design pattern in which a database table is modeled in terms of a class and a row of the database table is modeled by an instance of the class. The properties of the class correspond to the columns of the table. The ORM and the active-record pattern provide a consistent model and make it easier to access and manipulate the database from within the programming language. Another advantage of this approach is that programming time can be reduced by reusing many of the Castle ActiveRecord and NHibernate methods such as FindAll (find all records) or FindByProperty (find records related to a property such as all activities of a patient) to query for data without having to write Structured Query Language (SQL) queries.

Usability Testing

Usability of the system was assessed using the "think-aloud" approach [36-38]. In this approach, while participants are reviewing the system's content and interacting with the program, they are asked to vocalize thoughts, feelings, and opinions. The think-aloud approach gives an insight into how the user approaches the interface and what considerations the user keeps in mind when using the interface.

Think-aloud interviews were conducted with community providers (physicians and nurses, n = 3). A semistructured interview was used to collect input, and optional prompts were used if a provider did not continue to vocalize during the usability interview. The interview was conducted over the phone by study staff trained in the think-aloud protocols. Each interview was recorded and transcribed. Providers were asked to sign onto the ReferASmoker.org system, go through the registration process, and navigate through the site while making comments about their perceptions of the visual layout, as well as the location of options and functions within the system.

Phase 3: Implementation Planning

Once the primary processes were identified, we conducted an NGT session to collect feedback on the referral system and plan for implementation in practices. NGT is a highly structured, multistep, consensus-building procedure often used in formative research to elicit and prioritize group responses to a specific question. It is a consumer-oriented formal brainstorming or idea-generating technique used to foster creativity and to effectively prompt group members to articulate meaningful disclosures [39,40].

The study was conducted with a panel of experts (n = 9) that included health services researchers, and internal medicine and family practice providers. Using case scenarios, we introduced the goals of the study to the panel, as well as the proposed key components of the Web-based system identified in the process-mapping Delphi. The NGT sessions followed a standard protocol of solicitation of comments, discussion, and ranking of comments by level of importance. Questions posed were as follows: (1) What can we do to help you integrate ReferASmoker into your work clinic?, and (2) What would help you remember to use ReferASmoker?

Phase 4: Pilot Implementation and Evaluation

We tested implementation of the system to identify recruitment barriers and areas of refinement in the system. We recruited providers from family practice clinics to participate in the pilot study. Practices in the pilot were representative of participants in our planned larger trial. Using methods from a previously published randomized trial [41], we mailed 400 interest surveys that included a brief letter of introduction and a 1-page survey to determine a provider's interest and eligibility to participate in the project. Providers could respond to the interest survey online, by fax, or by mail using a prepaid, self-addressed envelope. If chosen for inclusion, providers were mailed a practice survey with a $150 incentive for completion.

Once the practice survey was completed, participants were mailed instructions on how to access and register on the website. We then measured the participant's usage of the system by tracking their interactions with the website. These data included the pages visited as well as the number of patient referrals on the system. After a period of use, each enrolled practice was contacted by telephone for follow-up; we assessed potential barriers and facilitators to future implementation at that time.

Results

Phase 1: System Conceptualization

We presented results of the literature review to our multidisciplinary research panel with expertise in health services, tobacco control, and informatics. Through the Delphi, our panel identified three key functionalities that would serve to overcome gaps in smoking-cessation referrals in clinical practices.

First, the research panel identified the importance of passive referrals such as information prescriptions in cessation efforts [42]. The panel recommended that providers use an information prescription approach with the ability to refer patients directly into an electronic system at the point of care. This Refer Your Smokers functionality would require a patient identifier, such as an email address, to be entered into a secure Web form or desktop client. Then, the system would automatically send active email reminders to patients encouraging participation.

Second, sustained cessation is difficult, and providers do not always have the benefit of observing the positive impact of increased counseling and referral activities. Their attention to

XSL•FO
RenderX

smoking cessation has little short-term positive reinforcement. In other referral processes for preventive care, there is often a *proximaloutcome*—a report of the result of screening. These reports (eg, results of a Pap test) produce a feedback loop and allow for an observable impact. Thus, our panel recommended creating practice reports that detail (1) the number of patients referred, and (2) the number of referred patients actually participating. These rates could be compared with other participating providers and potentially increase referrals.

Third, although many clinic-based interventions refer patients to public health services, such as quitlines, we noted almost no literature on referrals from public health interventions back into clinical care. Recent advances in prescription pharmacotherapy to aid smoking cessation make referral back to the provider for pharmacotherapy even more important. Thus, public health interventions should include content emphasizing the importance of seeking clinical treatment when a patient is ready to quit. The patient website should provide information about how to talk to your doctor about quitting and information about medications. For facilitating linkage back to clinical services, the panel recommended that patient and provider be connected via a secure messaging system. Thus, patients would be supported in the follow-up process, and providers could more easily assist with treatment and arrange follow-up.

In summary, based on the findings of the Delphi process, we conceptualized the following: (1) the system should support direct referral at point of care, (2) the system should provide continuous reports on patient activities to encourage continued participation of the providers, and (3) the system should support linkage of patients back to clinical services.

Additional functionalities were conceptualized to support the core functionalities noted above, including (1) a "quick-start" guide to train providers to use the system, (2) educational cases and materials to enhance provider knowledge about smoking cessation, (3) downloadable tools to support practice workflow (eg, posters to be used as cues for referral), and (4) methods for engaging providers longitudinally in the system (eg, a "headlines" section with evolving content, continuing education credit for educational cases, and an email reminder system to encourage referrals).

Phase 2: Programming and Usability testing

Website Functions

The ReferASmoker.org Web-based system was programmed using ASP.Net and C# technology (Figure 2). The following functions were developed: Refer Your Smokers, practice reports, secure messaging, and registration.

The core Refer Your Smokers function allows providers to proactively refer and enroll patients in the smoking-cessation system during the clinical encounter. To refer a patient, the provider logs into the ReferASmoker.org system and enters a willing patient's email address. Patients can be referred in multiples or one at a time. The patient referral triggers several automated processes: (1) the patient's email is entered into the database of the patient online smoking-cessation system, enabling the patient to register and login to the patient system, (2) the system links the patient with the appropriate practice and provider, enabling the practice reports and secure messaging functions, and (3) a series of automated emails is sent to encourage the patient to login to the smoking-cessation system.

The practice reports feature was specifically designed to increase observability of provider impact in supporting patients who smoke to take steps to improve their health by quitting (Figure 3). This function allows providers to monitor their patient smoking-cessation activities in real time. Several components of activity for providers are detailed, including (1) the numbers of patients referred, (2) the number of referred patients actually participating in the program, and (3) a comparison of these rates with other participating providers from practices across the country.

The secure messaging function was designed to enhance provider–patient communication. Providers can send messages to their patients to encourage use of the patient portal in their smoking-cessation efforts. For convenience, the system provides message templates, but providers have the option to customize them during their registration into the ReferASmoker.org system. A link to the secure messaging function is located within the ReferASmoker.org system so that providers have enhanced communication capabilities with their patients, who also receive this benefit on the portal. Providers can also initiate message threads within the secure messaging system.

In addition, we developed a toolbox of educational materials, interactive cases, and news headlines on the website. These materials were developed to supply providers with more general resources and materials to aid in the implementation of smoking-cessation strategies. The interactive cases were followed by questions testing comprehension of the information. Links were embedded in the interactive feedback to redirect the provider to different sections of the education materials to obtain additional information. On completion, providers earned 1 American Medical Association Physician's Recognition Award category 1 continuing medical education credit for each case. A registration process was created for a provider to register to the system using an email and password combination. The registration process included online consent, a survey, and two customizable email messages to the patient. Once the registration was completed, the provider could login to the system on the home page using the email password combination.

XSL•FO
RenderX

Figure 2. ReferASmoker.org Web-based system home page.

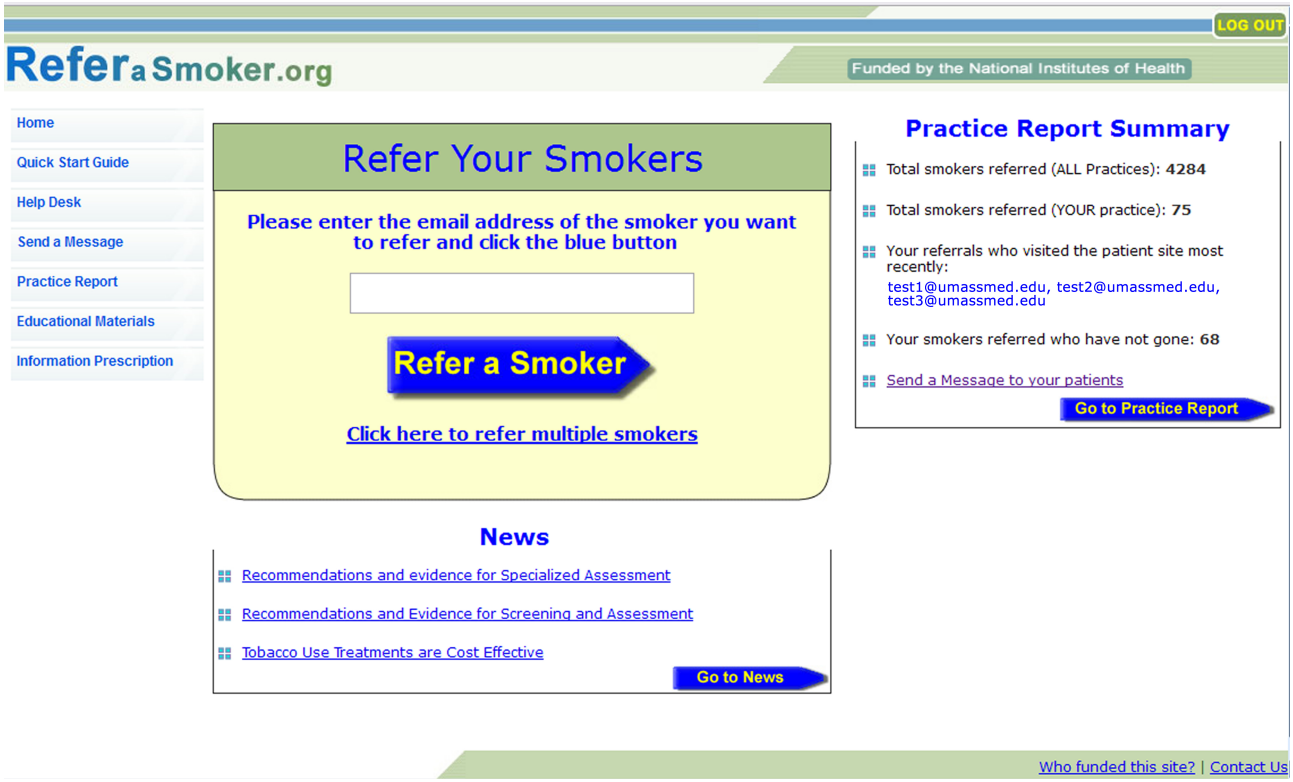
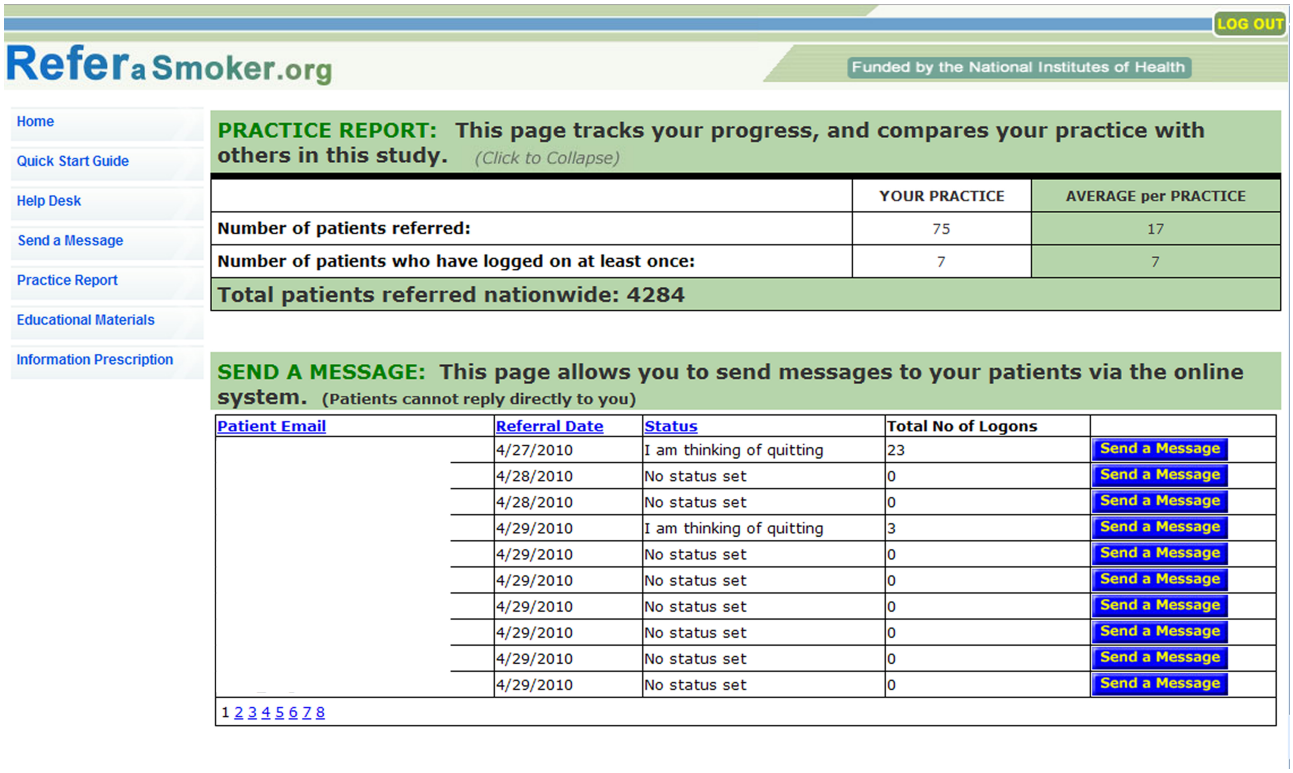


Figure 3. Sample ReferASmoker.org Web-based system practice report.



Web System Programming

Guided by the WCSF, we programmed the system using a modular and flexible architecture. We further divided the modules of the WCSF into the data access layer that enabled structured database access, the service layer that provided a collection of reusable functions, and the business process layer that orchestrated the functionality of system.

In the data access layer, using Castle ActiveRecord and NHibernate tools, we created ORM mappings between database tables in the SQL database to C# classes. The ORM mappings also included the relationships that exist between tables in the

XSL•FO
RenderX

SQL databases. For example, both a table for the list of providers and another table for the list of practices were developed. A many-to-one relationship exists between these tables (ie, a provider can belong to many practices and a practice can have many providers). The relationship was replicated in the ORM mappings using the BelongsTo and HasMany attributes of Castle ActiveRecord. The provider class possessed a property indicating that the provider BelongsTo the practice class and the practice class had a property indicating it HasMany providers.

In the services layer, we programmed "reusable" data query functions and common utilities that are used throughout the application. The reusable data query functions leverage the data access layer to perform query functions such as select, insert, and update. For example, the system contained provider functions that perform such operations as select all providers belonging to a practice, find the practice of the provider, or find the randomization of a particular practice. The utility functions included methods to send emails and encrypt and decrypt data. A SendEmails function was used throughout the system to send emails to patients, including transmission of secure emails when a provider uses the secure messaging function on the website or transmission of automated emails to encourage registration from the system. The provider and patient identifiers were stored in encrypted form in the database. For this, algorithms to encrypt and decrypt the provider and patients identifiers appropriately were programmed in the DataEncryption function.

In the business process layer, we programmed the business logic of the system—that is, a series of tasks that orchestrated the services to realize the functionality of the ReferASmoker.org processes, such as Refer Your Smokers, provider feedback, and secure messaging. For example, the Refer Your Smokers process performed several tasks that need to occur when a provider refers a patient, including (1) determining whether the patient was already referred in the database, (2) if a patient was already in the database, informing the provider that the patient was already referred, and (3) if it was a new patient, adding the patient's information (email, referring provider and referring practice information, referral date, and emails assigned for transmission from the provider to the patient) and informing the provider that the referral process was successful.

Usability Testing

Feedback acquired through the think-aloud usability testing was categorized into three themes: (1) registration and login process, (2) general layout, and (3) specific features. As the providers went through the registration process, several issues were identified. First, the instructions indicating that a new user must register and choose a password before using the system were not altogether clear. Second, the providers expressed displeasure with the system automatically assessing the strength of the password provided. Third, instructions for completing the registration survey and particular questions within the instrument were not clearly understood. Finally, the length of the registration process prevented completion of the usability testing process and was seen as a potential barrier to use of the system in practice.

Regarding the overall layout of the system, the providers indicated that the website was user-friendly and the various components self-explanatory. Providers expressed particular interest in the news headlines and education components of the system. Providers commented positively on the simplicity and ease of the Refer Your Smokers function. Providers were also pleased that once a referral was made, the system automatically emailed patients to remind them to visit the site. However, concern was expressed regarding the usefulness of the system for patients without email.

The practice reports were also believed to be of great utility. Providers remarked that the nationwide comparison of referrals and the real-time activity of their patient panel could serve as motivation to improve. The suggestions for improvement included adding an attention-getting visual to draw the eye to the status column, listing the most active patients at the top of the status report, and adding a mechanism that would announce a patient's first visit to the website or when a particular patient was doing very well or very poorly.

Providers were enthusiastic about the secure messaging function's potential to engage patients in their own care but provided several thoughts. First, the providers indicated value in the ability to print, download to an electronic medical record, or otherwise archive the messages sent for the medical records. Documenting these communications without additional work was seen as very important for proper follow-up and for possible reuse or modification in the future. Next, providers appreciated both the opportunity to use a preestablished message template that tailored content based on where the patient is in the quit process and the ability to make the messages more personal. Finally, providers commented that it might be beneficial for patients to have the ability to respond to the provider messages to engage them more in their care and in their quit processes, but they also acknowledged that a two-way communication path within the system could prove burdensome for many providers.

With regard to the educational toolbox, providers suggested that the various products should be labeled separately for convenience of location on the website. No matter how useful the information, busy clinicians would not spend precious time searching for the information. Further, more information for providers was suggested, including a quick-facts sheet with the latest statistics about smoking and links to the most relevant and recent evidence. Print options for all materials, including other treatment information (for patients), were also suggested.

Based on the feedback from the usability results, we made several changes to the system. An easily visible button with the text "NEW USERS! Please click this button to register" was created on the home page to clarify new-user registration instructions. The instructions on the survey page were also clarified. We removed the password strength feature on the username and password creation page. To reduce the additional step of logging in the system after registration, users were redirected to the home page after completing registration. In response to the comments on the practice reports, we created a practice report summary on the home page that contained the following information: (1) a count of the numbers of smokers

XSL•FO
RenderX

referred using the system by all practices and the number of smokers referred by the current practice, (2) emails of the last 3 patients of the current practice who were active on the patient site, and (3) the number of smokers of the current practice who have not visited the patient website. We did not create two-way secure messaging between the provider and patient because we felt that this will add additional burden to the providers. To improve the educational materials section, we further classified the materials into three sections: practice forms, interactive cases, and patient education. The first two sections grouped materials for increasing the knowledge and awareness of the provider. The latter sections, though delivered to the provider, contained materials for the provider to use for educating patients.

Phase 3: Implementation Planning

From the NGT session, we identified that several cues to action would be needed to implement ReferASmoker.org in practices, including workflow items and continuous reminders. First, NGT participants emphasized the importance of communicating with the practice through a contact person. This person would serve as the liaison with the practice over a set period of time to inquire about patient recruitment or any other questions or concerns with the system. Second, incentives for participation (e-referrals) were recommended. Third, a continuous communication plan, including both mail and email campaigns, was suggested for ongoing practice engagement. Participants indicated that regular emails would update participants about study progress and provide other information relevant to smoking cessation. Emails with embedded weblinks would provide convenient access back to the system. Fourth, in addressing practice workflow issues, it was recommended that hardcopy materials be sent to the practice to facilitate collection of patient email addresses and website instructions. Finally, NGT participants suggested that successes be appropriately celebrated, perhaps with emails of congratulations and gratitude to practices that logged into the study.

Phase 4: Pilot Implementation and Evaluation

In the pilot implementation, 25 practices out of 400 responded to a mailed survey indicating that they were interested in the project, and all of them were mailed a consent form. Of the 25 practices, 8 returned the consent form and were then mailed a practice survey. Of these 8 practices, 7 returned the survey and were given access to the ReferASmoker.org system. Out of those, 6 providers from 5 practices registered with the system, and 5 of them logged into the system. Initially, no providers referred patients. The principal investigator of the study contacted each of the practices by phone to encourage them to use the website. After the call, 1 provider used the referral function to refer 2 patients. Among these, 1 patient visited the patient website.

Telephone calls from the principal investigator to enrolled practices were not included in the original pilot implementation and evaluation protocol. However, it became important to elicit information from providers at this stage that could prove helpful in the main trial. We attempted phone contact with all 6 enrolled providers and succeeded in talking with 4. The providers reported barriers and facilitators to practice implementation. Overall, the providers liked the system and thought the intervention was a good idea, but had trouble implementing the system. The staff in the practices constantly changed and newer staff members were unaware of the study. Practices also did not remember whether they had registered with the system. Practices also forgot to e-refer because of lack of visual cues to the intervention. One provider summarized this succinctly, stating "I guess it's out of sight, out of mind." Providers encouraged cues to action, with suggestions for a waiting or examination room display that would serve as reminders to refer or to activate patients to talk to them about smoking cessation. Providers were not sure whether the system would be applicable to all patients. All providers agreed that an implementation budget would provide incentives for use of the system.

Discussion

In the preimplementation stage of a nationwide study of an interactive, Web-delivered system to increase provider and patient engagement in smoking cessation, we conducted a rigorous planning and evaluation of the system. The primary purpose of our preimplementation evaluation was to identify the strengths that might be used to promote the program, and weaknesses that might be mitigated prior to initiating the main study. We conceptualized and developed e-referral functions in Web-based form. We report the functions we developed and the results of our usability testing in the Results section. We evaluated the Web system and the implementation plan rigorously with community-based providers. Our approach involved four phases: (1) system conceptualization, (2) agile programming and think-aloud usability testing, (3) implementation planning (using the NGT), and (4) lessons learned from pilot implementation in 7 physician practices. Table 1 summarizes the identified barriers and facilitators to practice implementation based on our evaluation work. In the section below, we focus on the implementation protocol changes that will be used in the main trial to address the four primary barriers we uncovered in the pilot testing.

XSL•FO
RenderX

Table 1. Identified issues related to e-referral system implementation

Issue	Barrier	Facilitator	Identified by	Identification stage
Difficulty contacting the practice and lack of study champion	X		Study team	Pilot
Lack of training	X		Study users (providers)	Pilot
Registration difficulties	X		Study team and study users (providers)	Think-aloud usability
Lack of motivation and start-up incentives	X		Study team and study users (providers)	Think-aloud usability and pilot
Forgetting to refer	X		Study users	Pilot
Ease of system use		X	Study users (providers)	Think-aloud usability and pilot
Perceived potential to affect care		X	Study users (providers)	Think-aloud usability and pilot

The first barrier was the difficulty contacting the practice and lack of study champion. With no champion identified at each practice, we were constantly speaking with or leaving messages for different staff members, who had little sense of ownership of or urgency in the process. This breakdown in communication was made more complicated with staff turnover, a reality in most medical offices. In order to overcome this particular barrier, we modified the study protocol to include a request for each practice to identify two staff members to serve as implementation coordinators. These implementation coordinators will be the primary contacts for the practice and will work closely with our study personnel. Their responsibility will be to implement and promote the study intervention in the practice. Two implementation coordinators will allow for backup in the event that one individual is unavailable or leaves the practice. Our study personnel will communicate with these implementation coordinators to confirm practice information, hold training sessions, answer any questions, and provide feedback.

Second, we identified that successful implementation required training and assistance with registration in the system. Pilot practices reported that the system was easy to use, but with no one trained at the practice to complete the registration process and refer patients and to champion others through the process, the task went undone. Consequently, we increased study personnel and created a proactive helpdesk to provide training and help with registration. In the main study, our staff will initiate contact with each practice within 2 weeks of receiving the returned consent form. Study personnel will verify practice information and schedule a training and registration call for each of the implementation coordinators. During this call, our staff will walk the implementation coordinators through the actual registration process. The study personnel will be on hand to answer any questions. Following registration, study personnel will review the process for referring patients, getting the implementation coordinators to enter a dummy referral to have the full experience of the ease of referral. Each implementation coordinator will be encouraged and provided information to train others in the office to also register and refer patients. Following the training call, study personnel have planned a booster call to verify receipt of printed materials sent and answer any questions that may have arisen in the first few attempts to refer. If no referrals have been made yet, our staff will assess any reasons for no referrals and encourage implementation coordinators to use the system.

The third barrier was a lack of motivation and start-up incentives. It became abundantly clear that motivation to participate was low. We focused on increasing both extrinsic and intrinsic motivation. Pilot data immediately indicated that financial incentives would spur participation. Additional funds from the American Recovery and Reinvestment Act has enabled us to provide honoraria to participating practices and individual staff members who complete training and surveys.

Finally, we learned that clinicians within participating practices simply forgot about the study and the need to refer patients through the system. We believe that with convenient reminders we will be able to activate them to use the system. In addition to calling the implementation coordinators to aid them in the registration process and answer any questions, we will increase the work-flow support. We also improved the printed information prescription pads sent to practices for distribution to their patients simultaneously with their online referral. The "Information Rx" that was used in the pilot was small, about the size of a regular prescription pad, and simply provided an optional patient handout. The new and improved pad is spiral bound and has easy check-off boxes with duplicate pages. The bottom half of the first page will be given to the patient. The top half will be returned to study personnel, and the duplicate copy will be ready to place in a patient medical record file. Further, the information prescriptions for the intervention arm have a space for the providers to write the patient's email address. Additionally, posters to serve as visual stimulation to use the system, posters to encourage patients to talk with their provider about quitting, and 1-page instruction sheets outlining the steps for referring patients will be sent to participating practices.

To increase intrinsic motivation and to maximize the brief phone contact with practices, study personnel will incorporate a concept called motivational interviewing into each interaction. Miller and Rollnick define motivational interviewing as a client-centered, directive method for enhancing intrinsic

XSL•FO
RenderX

motivation to change by exploring and resolving ambivalence [43]. Key concepts involved in motivational interviewing are accurate empathy, reflective listening, and overcoming ambivalence, with the ultimate goal of facilitating some changed behavior. From our pilot study, we learned that implementation coordinators must face an increased workload because of their participation in the study enrollment, training, and implementation, especially in system registration and patient referral. Using motivational interviewing techniques, our staff will be better able to communicate effectively with implementation coordinators by identifying and overcoming their ambivalence. As an example, if a study staff member has attempted to contact a particular implementation coordinator on multiple times and failed, it may be a natural inclination to reflect negatively on that statement, which can be highly detrimental to the relationship with that particular office. Focusing on a positive reflective statement might increase the likelihood of the practice either becoming or continuing to be a happy participant in the study and increase positive feedback, which will, in turn, facilitate study task completion.

Results of Implementation Protocol Changes

Because of our implementation protocol changes, we were successful in engaging practices and improving participation in the nationwide trial. We measured rates of referral and patient participation in the first 3 months of practice engagement. To date, we have analyzed data from the first 50 e-referral practices. Practices' mean e-referral rate was 14 (SD 13.63). In the first 3 months, the maximum number of referrals by a practice was 62, and 3 practices did not refer. Per practice, the patients mean registration rate was 3.4 (SD 5.09). The maximum number of patients registered with a practice was 28, and 13 practices did not yet have any patients registering.

Strengths and Limitations

In preparation for a nationwide randomization trial testing an e-referral process for referring patients to a smoking-cessation system by providers, we detail the rigorous steps taken to develop the Web-based e-referral system. At each step of the development process, we applied user input to conceptualize and refine the system. Although the numbers of users are low, the multiple and comprehensive nature of the interactions and data collected provided significant information on which to improve the system's usability. The results of the mini pilot study gave us critical insight into the recruitment and use barriers that our randomized trial must overcome to succeed.

Conclusion

Our how-to report demonstrates how a small, rigorously conducted, multistep preimplementation evaluation can affect the success of a larger study. To gain valuable information regarding potential improvements to an interactive, Web-delivered provider–patient system to increase engagement in smoking cessation, we used a multidimensional approach to conceptualize, develop, implement, and test the product and process. The results of this rigorous process led us to make significant changes to the practice implementation approach study, prior to its nationwide randomized, controlled trial. After refining our information system with usability testing, we further uncovered serious barriers to implementation: lack of study champions within the practice, lack of training and assistance in use of the system, and lack of motivation to participate. We identified several improvements to address and made changes to the main study protocol before trial implementation. Our preliminary analysis with the first 50 practices using the system for 3 months demonstrates the preimplementation evaluation was successful in overcoming the barriers to recruiting and retain study participants.

Acknowledgments

For this project, the authors and manuscript preparation were supported by grant 5R01CA129091-04 from the National Cancer Institute.

Conflicts of Interest

None declared

References

1. Centers for Disease Control and Prevention (CDC). Cigarette smoking among adults--United States, 2000. MMWR Morb Mortal Wkly Rep 2002 Jul 26;51(29):642-645 [FREE Full text] [Medline: 12186222]
2. Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. Lancet 1997 May 24;349(9064):1498-1504. [doi: 10.1016/S0140-6736(96)07492-2] [Medline: 9167458]
3. Critchley JA, Capewell S. Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: a systematic review. JAMA 2003 Jul 2;290(1):86-97 [FREE Full text] [doi: 10.1001/jama.290.1.86] [Medline: 12837716]
4. Fiore MC, Croyle RT, Curry SJ, Cutler CM, Davis RM, Gordon C, et al. Preventing 3 million premature deaths and helping 5 million smokers quit: a national action plan for tobacco cessation. Am J Public Health 2004 Feb;94(2):205-210 [FREE Full text] [Medline: 14759928]
5. Fiore M. Treating Tobacco Use and Dependence: 2008 Update (Clinical Practice Guideline). Rockville, MD: Department of Health and Human Services, Public Health Service; 2008.

XSL•FO
RenderX

6. Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M, Comparative Risk Assessment collaborating group (Cancers). Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. Lancet 2005 Nov 19;366(9499):1784-1793. [doi: 10.1016/S0140-6736(05)67725-2] [Medline: 16298215]
7. Lancaster T, Stead LF. Self-help interventions for smoking cessation. Cochrane Database Syst Rev 2005(3):CD001118. [doi: 10.1002/14651858.CD001118.pub2] [Medline: 16034855]
8. Sciamanna CN, Ford DE, Flynn JA, Langford C. An evidence-based interactive computer program to assist physicians in counseling smokers to quit. MD Comput 1999;16(5):54-60. [Medline: 10570612]
9. Strecher VJ. Computer-tailored smoking cessation materials: a review and discussion. Patient Educ Couns 1999 Feb;36(2):107-117. [Medline: 10223016]
10. Bental DS, Cawsey A, Jones R. Patient information systems that tailor to the individual. Patient Educ Couns 1999 Feb;36(2):171-180. [Medline: 10223021]
11. Hersey JC, Matheson J, Lohr KN. US Department of Health and Human Services, Agency for Health Care Policy and Research. 1997 Sep. Consumer Health Informatics and Patient Decision-Making: Final Report URL: http://courseweb.edteched.uottawa.ca/nsg6533/rtifinal.pdf [accessed 2011-07-07] [WebCite Cache ID 600OmLk0Q]
12. Ramelson HZ, Friedman RH, Ockene JK. An automated telephone-based smoking cessation education and counseling system. Patient Educ Couns 1999 Feb;36(2):131-144. [Medline: 10223018]
13. Ossip-Klein DJ, McIntosh S. Quitlines in North America: evidence base and applications. Am J Med Sci 2003 Oct;326(4):201-205. [Medline: 14557735]
14. Anonymous. HHS 'Quitline' helps Americans stop smoking. FDA Consum 2005;39(1):4-5 [FREE Full text] [Medline: 15818793]
15. Velicer WF, Prochaska JO. An expert system intervention for smoking cessation. Patient Educ Couns 1999 Feb;36(2):119-129. [Medline: 10223017]
16. Patten CA, Smith CM, Brockman TA, Decker PA, Anderson KJ, Hughes CA, et al. Support person intervention to promote smoker utilization of the QUITPLAN Helpline. Am J Prev Med 2008 Dec;35(6 Suppl):S479-S485. [doi: 10.1016/j.amepre.2008.09.003] [Medline: 19012842]
17. Miller CL, Wakefield M, Roberts L. Uptake and effectiveness of the Australian telephone Quitline service in the context of a mass media campaign. Tob Control 2003 Sep;12 Suppl 2:ii53-ii58 [FREE Full text] [Medline: 12878774]
18. Fiore M. Treating Tobacco Use and Dependence. Washington, DC: US Department of Health and Human Services, Public Health Service; 2000.
19. Shelley D, Cantrell J. The effect of linking community health centers to a state-level smoker's quitline on rates of cessation assistance. BMC Health Serv Res 2010;10:25 [FREE Full text] [doi: 10.1186/1472-6963-10-25] [Medline: 20100348]
20. Holtrop JS, Malouin R, Weismantel D, Wadland WC. Clinician perceptions of factors influencing referrals to a smoking cessation program. BMC Fam Pract 2008;9:18 [FREE Full text] [doi: 10.1186/1471-2296-9-18] [Medline: 18373854]
21. Godfrey MM, Nelson EC, Wasson JH, Mohr JJ, Batalden PB. Microsystems in health care: Part 3. Planning patient-centered services. Jt Comm J Qual Saf 2003 Apr;29(4):159-170. [Medline: 12698806]
22. Huber TP, Godfrey MM, Nelson EC, Mohr JJ, Campbell C, Batalden PB. Microsystems in health care: Part 8. Developing people and improving work life: what front-line staff told us. Jt Comm J Qual Saf 2003 Oct;29(10):512-522. [Medline: 14567260]
23. Wasson JH, Godfrey MM, Nelson EC, Mohr JJ, Batalden PB. Microsystems in health care: Part 4. Planning patient-centered care. Jt Comm J Qual Saf 2003 May;29(5):227-237. [Medline: 12751303]
24. Houston TK, Sadasivam RS, Ford DE, Richman J, Ray MN, Allison JJ. The QUIT-PRIMO provider-patient Internet-delivered smoking cessation referral intervention: a cluster-randomized comparative effectiveness trial: study protocol. Implement Sci 2010;5:87 [FREE Full text] [doi: 10.1186/1748-5908-5-87] [Medline: 21080972]
25. Adler M, Ziglio E. Gazing into the Oracle: The Delphi Method and Its Application to Social Policy and Public Health. London: Jessica Kingsley Publishers; 1996.
26. Linstone HA, Turoff M. The Delphi Method: Techniques and Applications. Reading, MA: Addison-Wesley, Advanced Book Program; 1975.
27. Agile Programming. 2010. Agile & Scrum Programming URL: http://agileprogramming.org/ [accessed 2010-12-23] [WebCite Cache ID 5vC4wxWkM]
28. Gamma E. Design Patterns: Elements of Reusable Object-Oriented Software. Reading, MA: Addison-Wesley; 1995.
29. p&p Client Team. Microsoft patterns & practices. 2010 Jun 2. Web Control Software Factory URL: http://webclientguidance.codeplex.com/ [accessed 2011-06-15] [WebCite Cache ID 5zT9qiBF2]
30. Wikipedia. 2010 Jun 4. Model-view-presenter URL: http://en.wikipedia.org/w/index.php?title=Model-view-presenter&oldid=366065436 [WebCite Cache ID 5zT9tu91I]
31. Patricio A. Hibernate. 2010. NHibernate for .NET URL: http://community.jboss.org/wiki/NHibernateForNET [accessed 2011-06-15] [WebCite Cache ID 5zTA0vwnS]
32. Castleproject. Castle ActiveRecord URL: http://www.castleproject.org/activerecord/index.html [accessed 2011-06-15] [WebCite Cache ID 5zTA4bZkJ]

XSL•FO
RenderX

33. Ferraggine VE, Doorn JH, Rivero LC. Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends. Hershey, PA: Information Science Reference; 2009.
34. Mitchell S. MSDN Magazine. 2009. The Active Record Pattern, Aspect-Oriented Programming URL: http://msdn.microsoft.com/en-us/magazine/dd419658.aspx [accessed 2011-06-15] [WebCite Cache ID 5zTA88xXI]
35. Fowler M. Patterns of Enterprise Application Architecture. Boston, MA: Addison-Wesley; 2003.
36. Crowley RS, Naus GJ, Stewart J, Friedman CP. Development of visual diagnostic expertise in pathology -- an information-processing study. J Am Med Inform Assoc 2003;10(1):39-51 [FREE Full text] [Medline: 12509356]
37. Kushniruk AW. Analysis of complex decision-making processes in health care: cognitive approaches to health informatics. J Biomed Inform 2001 Oct;34(5):365-376. [doi: 10.1006/jbin.2001.1021] [Medline: 12123153]
38. Kushniruk AW, Patel VL. Cognitive computer-based video analysis: its application in assessing the usability of medical systems. Medinfo 1995;8 Pt 2:1566-1569. [Medline: 8591502]
39. Grant A, Berlin A, Freeman GK. The impact of a student learning journal: a two-stage evaluation using the Nominal Group Technique. Med Teach 2003 Nov;25(6):659-661. [doi: 10.1080/01421590310001605714] [Medline: 15369916]
40. Miller D, Shewchuk R, Elliot TR, Richards S. Nominal group technique: a process for identifying diabetes self-care issues among patients and caregivers. Diabetes Educ 2000;26(2):305-10, 312, 314. [Medline: 10865596]
41. Houston TK, Richman JS, Ray MN, Allison JJ, Gilbert GH, Shewchuk RM, DPBRN Collaborative Group. Internet delivered support for tobacco control in dental practice: randomized controlled trial. J Med Internet Res 2008;10(5):e38 [FREE Full text] [doi: 10.2196/jmir.1095] [Medline: 18984559]
42. Bentz CJ, Bayley KB, Bonin KE, Fleming L, Hollis JF, McAfee T. The feasibility of connecting physician offices to a state-level tobacco quit line. Am J Prev Med 2006 Jan;30(1):31-37. [doi: 10.1016/j.amepre.2005.08.043] [Medline: 16414421]
43. Miller WR, Rollnick S. Motivational Interviewing: Preparing People for Change. 2nd edition. New York, NY: Guilford Press; 2002.

Abbreviations

MVP: model-view-presenter
NGT: nominal group technique
ORM: object-relational mapping
SQL: Structured Query Language
WCSF: Web Client Software Factory

Edited by G Eysenbach; submitted 28.12.10; peer-reviewed by J Gordon, E Smit; comments to author 19.04.11; revised version received 17.06.11; accepted 05.07.11; published 18.10.11.

Please cite as:
Sadasivam RS, Delaughter K, Crenshaw K, Sobko HJ, Williams JH, Coley HL, Ray MN, Ford DE, Allison JJ, Houston TK
Development of an Interactive, Web-Delivered System to Increase Provider–Patient Engagement in Smoking Cessation
J Med Internet Res 2011;13(4):e87
URL: http://www.jmir.org/2011/4/e87/
doi:10.2196/jmir.1721
PMID:22011394

©Rajani S Sadasivam, Kathryn Delaugther, Katie Crenshaw, Heather J Sobko, Jessica H Williams, Heather L Coley, Midge N Ray, Daniel E Ford, Jeroan J Allison, Thomas K Houston. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 18.10.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health Promotion in the Workplace: Assessing Stress and Lifestyle With an Intranet Tool

Daniela Lucini[1], MD, PhD; Nadia Solaro[2], BS; Alessandro Lesma[3], MD; Veronique Bernadette Gillet[3], BS; Massimo Pagani[1,4], MD

[1]Centro di ricerca sulla Terapia Neurovegetativa e Medicina dell'Esercizio, Dipartimento Scienze Cliniche, Università degli Studi di Milano, Milan, Italy

[2]Dipartimento di Statistica, Università degli Studi di Milano-Bicocca, Milan, Italy

[3]ENI corporate, Milan, Italy

[4]U.O. Telemedicina e Medicina dello Sport, Ospedale "Luigi Sacco", Milan, Italy

Corresponding Author:
Daniela Lucini, MD, PhD
Centro di ricerca sulla Terapia Neurovegetativa e Medicina dell'Esercizio
Dipartimento Scienze Cliniche
Università degli Studi di Milano
Via G B Grassi 74
Milan, 20157
Italy
Phone: 39 0239042748
Fax: 39 2050319823
Email: daniela.lucini@ctnv.unimi.it

Abstract

Background: Chronic noncommunicable conditions, particularly cardiovascular and metabolic diseases, are the major causes of death and morbidity in both industrialized and low- to middle-income countries. Recent epidemiological investigations suggest that management of lifestyle factors, such as stress and lack of physical activity, could have an important value in cardiometabolic conditions, while information technology tools could play a significant facilitatory role.

Objectives: The objective of our study was to verify the feasibility of using a private website, directed to the workers of a major Italian company, to describe their health profile and lifestyle and work habits using an ad hoc self-administered questionnaire.

Methods: We administered anonymous multiple choice Web-based questionnaires to 945 participants (683 completed the task) as part of an ongoing health promotion program in a multinational company. Qualitative and quantitative data were synthesized with nonlinear principal component analysis to construct indicators (ie, variables) for stress, control, and lifestyle domains. Considering in addition absenteeism, the Calinski-Harabasz statistic and cluster analysis jointly differentiated seven clusters, which displayed different distributions of standardized classification variables. The final step consisted in assessing the relationship of the resulting seven subject typologies with personal data, illnesses, and metabolic syndrome status, carried out for the most part with descriptive methods.

Results: Statistical analyses singled out two not-overlapping domains of stress and control, as well as three not-overlapping domains of physical activity, smoking, and alcohol habits. The centroids of the seven clusters generated by the procedure were significantly ($P < .001$) different considering all possible 21 comparisons between couples of groups. Percentage distributions of variables describing personal information (gender, age group, work category, illness status, or metabolic syndrome) within participant typologies show some noteworthy findings: females, workers aged 35–44 years, junior white collar workers, and respondents reporting illness were more prevalent in the stress group than in the overall studied population; preclinical metabolic syndrome status was more prevalent in the group with higher alcohol consumption. Absentees reported more illness.

Conclusions: The present Intranet-based study shows the potential of applying diverse statistical techniques to deal jointly with qualitative and quantitative self-reported data. The resulting formal description of subject typologies and their relationship with personal characteristics might provide a convenient tool for supporting health promotion in the work environment.

(J Med Internet Res 2011;13(4):e88) doi:10.2196/jmir.1798

XSL•FO
RenderX

KEYWORDS

Stress; lifestyle; risk factor; prevention; web-based questionnaire

Introduction

Background

Chronic noncommunicable conditions, particularly cardiovascular and metabolic diseases, are the major causes of death and morbidity around the world, affecting both industrialized and low- to middle-income countries. Recent epidemiological investigations [1,2] showed that lifestyle factors, such as stress and lack of physical activity, provide additional prognostic information to that furnished by usual coronary risk factors, suggesting that their management might have clinical value [3-5]. Behavioral components of risk are, in addition, difficult to handle because they cannot be treated by traditional pharmacological means and require the active collaboration of patients, who must change their attitudes and habits [6-9]. Lifestyle components, such as stress and exercise, have the advantage of being assessable individually with information technology (IT)-administered questionnaires [10], although their self-reported nature mandates additional caution in interpreting findings [11]. Stress can be described by various personal (symptom profile, psychological distress, and fatigue perception), social (family and work environment) [12], and functional domains (autonomic and hormonal regulation) [12-16]. Physical exercise can be defined in terms of intensity, modality (strength or aerobic training), and duration, leading to an algorithmic evaluation of workout level in a given time period [9,17,18].

Maintaining an ideal health risk profile in middle age might have important implications for greater longevity, compression of disease, increased quality of life, and reduced costs [19]. Because only a very limited fraction of the population (about 5%) fits into the ideal risk limits, new techniques must be tested to reach these new goals; these techniques might encompass the introduction of Web applications [20] with a focus on lifestyle [21]. We have been testing a behavioral approach to cardiovascular prevention, focusing on stress and inactivity in addition to usual risk factors, in various settings ranging from the outpatient clinic [15] to the workplace [13]. We also tested the feasibility of simple IT applications for technician-mediated [22] or self-reported [10] data entry. The use of Web- (or Intranet-) based approaches might also be suitable to deliver essential training with digital techniques and minimal cost [22-26], accommodating any personal preference for site, time, or pace, possibly also furnishing useful clinical feedback, whereby congruent multiparameter models, such as metabolic syndrome (MeS) [27], might be easier to handle [22].

Aims

With this in mind we designed this exploratory investigation to verify the feasibility of using an Intranet-based tool in the workplace [10] intended to assist employees of an Italian company to optimize lifestyle and cardiometabolic risk, as part of a company's health promotion initiative. Specific constraints were strict anonymity, minimal investments, and specific targets of physical activity, eating habits, and stress, with adherence to the company's privacy policy. In this report we present a methodology to describe the baseline status of a group of employees who agreed to participate in this initial exploratory phase of the study. Metabolic risk was approximated by using the MeS model, according to the Adult Treatment Panel III (ATP III) definition [27].

Methods

This study is part of an ongoing Web-based health promotion initiative of a major Italian multinational company. At this initial stage of the project, through the company Intranet, workers were offered an information service on health promotion, focusing on various work- and non-work-related issues, ranging from influenza epidemics to healthy lifestyle. In addition, as a company benefit, workers could log on to the health portal and enroll in an educational project based on a self-administered Web questionnaire [10], eventually aiming at optimizing lifestyle and minimizing cardiometabolic risk. According to the company's policy, the project required strict anonymity that was guaranteed by the use of name and password protection chosen by participants and maintained secret.

The Health Questionnaire

The anonymous questionnaire, which is an extended project-tailored version of the Subjective Stress Symptoms Scores Questionnaire (4SQ; previously described [13-16]), contains 50 multiple choice questions, addressing various domains related to working conditions (job level and absenteeism), living and exercise habits, and perceived stress and control. In addition, participants were asked to gather (if available) their most recent (<3 months old) reports on blood chemistry (total and high-density lipoprotein cholesterol, triglycerides, glucose), blood pressure, heart rate, and anthropometrics (including waist circumference), and to enter these numerical values. Detailed instructions about compilation were given through a short movie tutorial available through the company's portal.

The number of questions was a balance between the time required to fill out the questionnaire and the detail of the inferences that could be drawn.

Weekly activity levels were estimated from the approximate daily amount (in minutes) and nominal intensity, and expressed in (estimated) metabolic equivalents (METs), using a validated approach [17,18].

Psychological distress was estimated from the following items: bodily symptom perception, stress perception, fatigue perception, and control perception, as in our previous studies on this topic [13-16]. After answering the questionnaire, participants were provided with a graphical answer that illustrated potential areas of lifestyle improvement [10]. If participants subsequently wished to verify any changes possibly related to effects of lifestyle interventions, they were allowed to use the questionnaire again. The present study addressed only the initial descriptive part.

XSL•FO
RenderX

Participants

Participation in the study was under the guarantee of strict anonymity, and the questionnaire was made available for a month, from October to November 2008. At the time of the study about 24,000 workers of the Italian branch of the company had access to the Intranet as a fundamental instrument for everyday work. About 9970 workers accessed the complementary health portal every month; of these 4877 read information regarding the ongoing preventive campaign, 1380 saw the detailed instructions to fill out the questionnaire, and 945 employees elected to actively participate in the survey, on a completely voluntary basis. Employees were motivated to accurately fill out the questionnaire by the possibility of immediately obtaining an individual map of their risk factor profile and areas for improvement based on input data [10].

To optimize data quality, we excluded those participants who did not complete their reports or who provided unrealistic data, particularly regarding biochemistry, blood pressure, or anthropometrics. The final data set comprised 683 participants.

The protocol of the study was approved by the Institutional Ethics Committee, as part of an ongoing investigation on lifestyle-based prevention.

Statistics and Data Handling

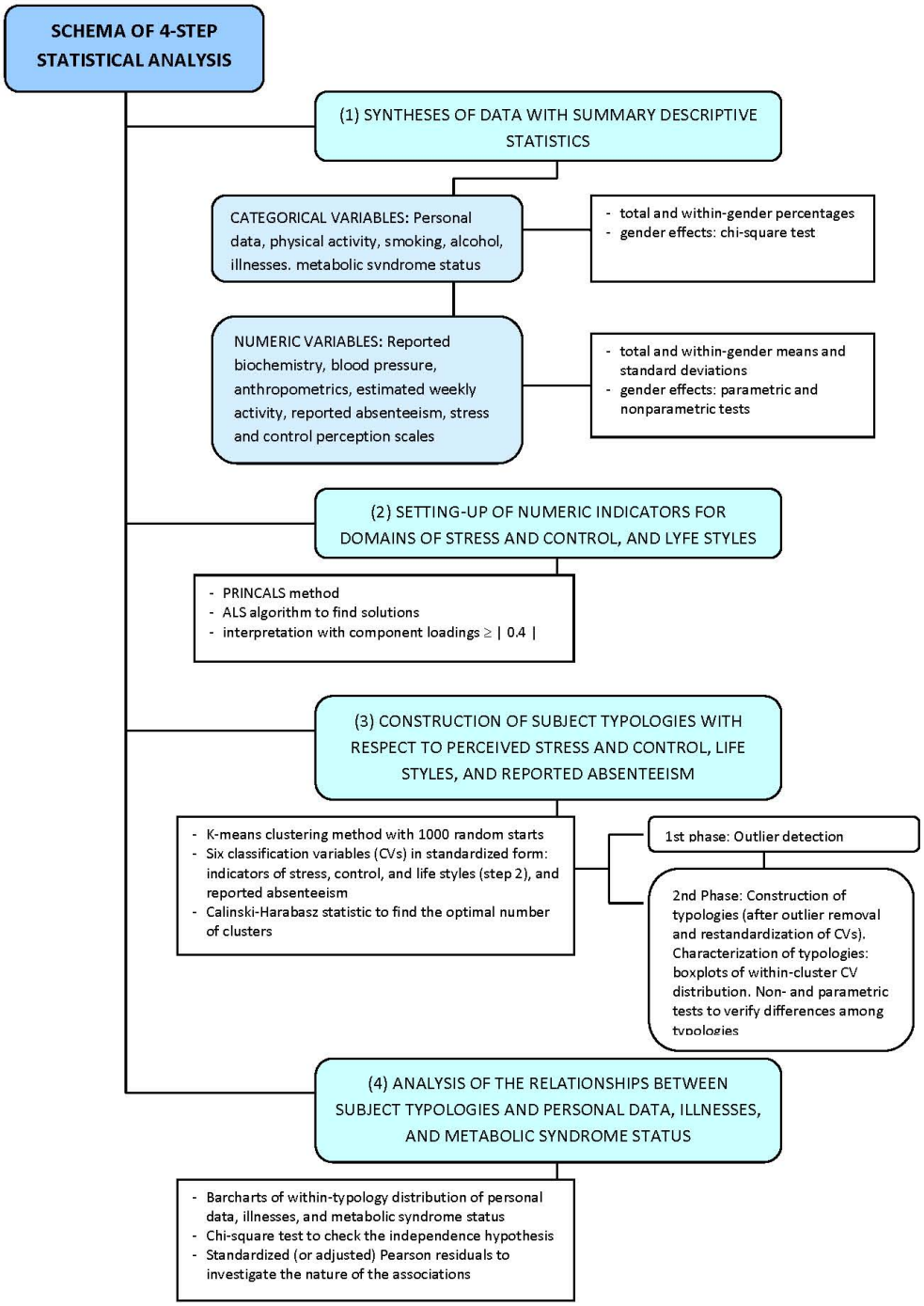
The main goal of disclosing and assessing possible relations linking cardiometabolic risk factors with perceived stress and control and with lifestyle is accomplished in four key steps (Figure 1): (1) synthesizing the information collected by the health questionnaire with summary descriptive statistics, (2) setting up quantitative indicators for perceived stress and control and for lifestyles, (3) building respondents' typologies with respect to perceived stress and control, lifestyles, and reported absenteeism, (4) assessing the presence of relationships between respondents' typologies and their personal data, illness status, and the MeS [27]. We mostly performed statistical analyses according to a data-driven approach by using exploratory multivariate data analysis techniques—that is, the nonlinear principal component analysis (PRINCALS) method [28] and k-means clustering algorithm [29]. We also performed statistical tests, although in this investigation inferences should have a minor role. The target population, to which inferences should be referred, was not clearly identified due to respondents' self-selection, thus suggesting some caution in our interpretations. We applied both parametric and nonparametric testing procedures [29,30] to take advantage of their specific potentials and to overcome their specific limits. We considered a test result to be "sufficiently revealing" if it was borne out as significant by both procedures.

With specific reference to the first step of analysis, summary descriptive statistics are presented in Table 1 and Table 2. The fourth column in Table 1 reports percentage distributions of personal data, lifestyle, and illness variables for the 683 participants in the study. To show the extent of gender differences, within-gender percentage distributions—percentages computed for each variable given (or conditionally to) the gender—are also provided in different columns. We performed a chi-square test to verify whether the above variables and gender could be assumed to be statistically independent (last column of Table 1).

Table 2 presents means and standard deviations of variables pertaining to reported biochemistry and anthropometrics, estimated weekly activity, reported absenteeism, and perceived stress and control scales. We also computed summary statistics for males and females separately. The presence of significant gender effects was assessed through both parametric univariate analysis of variance (ANOVA) and nonparametric Mann-Whitney and 2-sample Kolmogorov-Smirnov testing procedures. The (null) hypotheses to check for each variable were the equality of within-gender means (ANOVA) and the equality of the two within-gender distributions (Mann-Whitney and Kolmogorov-Smirnov).

XSL•FO
RenderX

Figure 1. Schematic outline of the phases of data analysis (ALS = alternating least squares, PRINCALS = nonlinear principal component analysis).



XSL•FO
RenderX

Table 1. Distribution of data (N = 683 participants): total and within-gender percentages

Variable	Male	Female	Total	*P* value[a]
Personal data				
Gender	495/683, 72.5%	188/683, 27.5%		
Work category				<.001
Blue collar	24/495, 4.8%	1/188, 0.5%	25/683, 3.7%	
Junior white collar	249/495, 50.3%	124/188, 66.0%	373/683, 54.6%	
Senior white collar	197/495, 39.8%	59/188, 31.4%	256/683, 37.5%	
Manager	25/495, 5.1%	4/188, 2.1%	29/683, 4.2%	
Age group (years)				ns[b]
<35	55/495, 11.1%	26/188, 13.8%	81/683, 11.9%	
35–44	133/495, 26.9%	58/188, 30.9%	191/683, 28.0%	
45–54	234/495, 47.3%	87/188, 46.3%	321/683, 47.0%	
>54	73/495, 14.7%	17/188, 9.0%	90/683, 13.2%	
Lifestyle				
Smoking habit				ns[b]
Nonsmoker	341/495, 68.9%	140/188, 74.5%	481/683, 70.4%	
Quit >1 year	58/495, 11.7%	20/188, 10.6%	78/683, 11.4%	
Quit <1 year	10/495, 2.0%	5/188, 2.7%	15/683, 2.2%	
≤5 cigarettes/day	22/495, 4.4%	8/188, 4.3%	30/683, 4.4%	
>5 and ≤10 /day	26/495, 5.3%	9/188, 4.8%	35/683, 5.1%	
>10 and ≤20/day	29/495, 5.9%	5/188, 2.7%	34/683, 5.0%	
>20/day	9/495, 1.8%	1/188, 0.5%	10/683, 1.5%	
Intend to quit				ns[b]
Yes, now	40/495, 8.1%	7/188, 3.7%	47/683, 6.9%	
Yes, in 6 months	24/495, 4.8%	8/188, 4.3%	32/683, 4.7%	
Does not intend to quit	30/495, 6.1%	8/188, 4.3%	38/683, 5.6%	
Structured physical activity				ns[b]
None	70/495, 14.1%	24/188, 12.8%	94/683, 13.8%	
No, but would like to	133/495, 26.9%	43/188, 22.9%	176/683, 25.8%	
Sometimes	51/495, 10.3%	21/188, 11.2%	72/683, 10.5%	
About 1 hour/week	64/495, 12.9%	32/188, 17.0%	96/683, 14.1%	
≤30 minutes/day, 3 times/week	81/495, 16.4%,	42/188, 22.3%	123/683, 18%	
≤30 minutes/day, 5 times/week moderate activity or ≤20 minutes/day, 3 times/week vigorous activity	60/495, 12.1%	16/188, 8.5%	76/683, 11.1%	
≤30 minutes/day every day moderate or intense activity	36/495, 7.3%	10/188, 5.3%	46/683, 6.7%	
Wine or beer (glasses/week)				<.001
None	72/495, 14.5%	78/188, 41.5%	150/683, 22%	
1–2	147/495, 29.7%	77/188, 41.7%	224/683, 32.8%	
3–7	169/495, 34.1%	21/188, 11.2%	190/683, 27.8%	
8–14	77/495, 15.6%	8/188, 4.3%	85/683, 12.4%	
15–21	20/495, 4.0%	3/188, 1.6%	23/683, 3.4%	
22–30	7/495, 1.4%	0/188, 0%	7/683, 1.0%	

XSL•FO
RenderX

Variable	Male	Female	Total	*P* value[a]
>30	3/495, 0.6%	1/188, 0.5%	4/683, 0.6%	
Alcohol (glasses/week)				<.001
None	361/495, 72.9%	178/188, 94.7%	539/683, 78.9%	
1–2	118/495, 23.8%	9/188, 4.8%	127/683, 18.6%	
3–7	15/495, 3.0%	0/188, 0.0%	15/683, 2.2%	
8–14	1/495, 0.2%	1/188, 0.5%	2/683, 0.3%	
Illnesses				ns[b]
None	316/495, 63.8%	116/188, 61.7%	432/683, 63.3%	
Functional illness	51/495, 10.3%	31/188, 16.5%	82/683, 12%	
Organic illness	128/495, 25.9%	41/188, 21.8%	169/683, 24.7%	
Metabolic syndrome status[c]				<.001
Normal	55/495, 11.1%	140/188, 74.5%	195/683, 28.6%	
Preclinical	285/495, 57.6%	46/188, 24.5%	331/683, 48.5%	
Metabolic syndrome	155/495, 31.3%	2/188, 1.1%	157/683, 23.0%	

[a] Significance level in the chi-square test for testing the null hypothesis of independence of variables and gender.

[b] Not significant ($P > .05$).

[c] Metabolic syndrome is inferred from data presented in Table 2.

XSL•FO
RenderX

Table 2. Descriptive data (N = 683 participants)[a]

Variables	Total		Male		Female		Reference values
	Mean	SD	Mean	SD	Mean	SD	
Reported biochemistry, blood pressure, and anthropometrics							
Total cholesterol (mg/dL)	203.23	38.03	203.92	38.76	201.43	36.07	<200
HDL[b] cholesterol (mg/dL)**,††,‡‡	60.01	23.37	56.68	23.13	68.79	21.70	Male: >29, female: >35
LDL[c] cholesterol (mg/dL)*,†,‡‡ (Friedewald formula)	120.48	40.08	122.61	41.49	114.86	35.59	<100
Triglycerides (mg/dL)**,††,‡‡	113.72	71.03	123.16	74.44	88.88	53.91	<150
Glucose (mg/dL)**,††,‡‡	90.09	16.63	91.89	17.79	85.34	11.90	74–106
Systolic blood pressure (mmHg)**,††,‡‡	122.73	11.23	124.51	10.17	118.06	12.50	<120
Diastolic blood pressure (mmHg)**,††,‡‡	78.54	7.91	79.44	7.20	76.15	9.12	<80
Heart rate (beats/minute)**,††,‡	70.09	10.12	69.22	10.15	72.37	9.70	60–90
Weight (kg)**,††,‡‡	75.27	13.60	79.72	11.57	63.55	11.43	NA[d]
Height (cm)**,††,‡‡	172.93	7.86	176.02	6.00	164.78	6.18	NA
Body mass index (kg/m^2)**,††,‡‡	25.06	3.61	25.71	3.39	23.35	3.62	<25
Waist circumference (cm)**,††,‡‡	90.42	12.06	93.76	10.30	81.63	11.98	Male: <102, female: <88
Estimated weekly activity (metabolic equivalents, minutes/week)							
Walking	436.17	451.37	455.98	469.64	384.03	395.75	
Moderate activity	378.38	445.24	370.26	449.14	399.73	435.27	
Vigorous activity**,††,‡‡	551.59	822.88	630.80	857.98	343.02	681.59	
Total activity**,†,‡	1366.14	1239.34	1457.05	1275.36	1126.78	1106.97	
Reported absenteeism							
Lost working days (in previous 12 months)††,‡‡	5.87	14.80	5.34	16.29	7.25	9.74	
Stress and control perception scales (AU)[e]							
4SQ[f] **,††,‡‡	16.86	20.11	14.75	18.53	22.41	22.91	
Stress**,††,‡‡	2.64	2.70	2.53	2.72	3.67	3.05	
Fatigue**,††,‡‡	2.84	2.86	2.28	2.54	3.60	2.88	
Control	4.11	3.16	4.17	3.27	3.95	2.83	

[a] Although in the subsequent steps of analysis statistical evaluation of perceived stress and control scales is performed in nonmetric terms, in this table, for practical reasons, they are presented as means and SD.

[b] High-density lipoprotein.

[c] Low-density lipoprotein.

[d] Not applicable.

[e] Arbitrary units.

[f] Subjective Stress Symptoms Score Questionnaire.

Significance level in the univariate analysis of variance (the null hypothesis is the equality of within-gender means): *significant at the .05 level, **significant at the .001 level. Actual *P* value for LDL cholesterol is *P* = .02.

Significance level in Mann-Whitney test (the null hypothesis is the equality of within-gender distributions): †significant at the .05 level, ††significant at the .001 level. Actual *P* value for LDL cholesterol is *P* = .01.

Significance level in Kolmogorov-Smirnov test (the null hypothesis is the equality of within-gender distributions): ‡significant at the .05 level, ‡‡significant at the .001 level. Actual *P* value for LDL cholesterol is *P* = .03; for heart rate is *P* = .01; for total activity is *P* = .01.

Regarding step (2), we set up numeric indicators to represent the underlying domains of stress, control, and lifestyles, with the categorical principal component analysis method (CATPCA, SPSS version 18; IBM Corporation, Somers, NY, USA), also

XSL•FO
RenderX

known in the statistical literature as nonlinear principal component analysis, or PRINCALS [28]. PRINCALS is an advanced multivariate statistical analysis technique addressed to data dimensionality reduction problems. It is still relatively little used in many fields of application, probably due to its intrinsic complexity, but it offers potentials in data analysis not shared by ordinary methods. Since it is not so well known yet, we provide a brief description of its main characteristics.

Unlike other methods such as principal component analysis (PCA) or factor analysis, PRINCALS can jointly handle qualitative and quantitative variables to convey their informational content in a small, a priori fixed number of dimensions (ie, unobservable variables), thus synthesizing data with the least possible loss of information. PRINCALS constructs dimensions that have zero mean and are pairwise uncorrelated by minimizing a loss function under several restrictions [28]. The minimization is made with respect to a set of unknown quantities, namely object (or dimension) scores (ie, values of dimensions) and category quantifications (ie, values that are attributed to categories of the original qualitative variables). Unlike PCA, the minimization problem does not admit a closed-form solution. It requires iterative procedures (Figure 1), specifically the alternating least squares algorithm. This is essentially a 2-step process that alternates updated solutions over object scores and category quantifications. Within the PRINCALS routine, qualitative variables are quantified, or optimally scaled, in the sense that their categories are replaced by metric values. So, optimal scaling transformations account for the different measurement level of variables. Nominal variables are quantified through application of the so-called centroid principle [28]. Ordinal variables are quantified through the weighted monotonic regression transformation, which allows the rank order to be preserved among ordinal categories. Numerical variables, being already metric, are generally simply standardized. More complex transformations can also be involved—for example, spline functions could be applied to nominal and ordinal variables [28]. Summing up, PRINCALS reaches two goals simultaneously: quantifying where necessary, and reducing the number of original variables (both qualitative and quantitative) by extracting dimensions. In addition, the PRINCALS routine automatically rotates the extracted dimensions to refer them to their principal axes in analogy with PCA. Accordingly, the computed dimensions reproduce the maximum possible variation in data or, more precisely, in optimally scaled variables.

After extraction, dimensions, being unobservable variables, require interpretation in order to establish which semantic fields or domains of original variables they account for. Interpretation is mainly based on the so-called component loadings, which are linear correlation coefficients of dimensions and optimally scaled variables. Dimensions assume the meaning from the variables with which they are more highly correlated, in a negative or positive sense. Usually, a threshold (absolute) value of 0.4 is introduced to distinguish negligible ($<|0.4|$) from essential ($\geq|0.4|$) component loadings. Once their meaning is assessed, dimensions are likely to be treated as indicators of the specific semantic fields they represent.

Another aspect of concern regards the choice of the "ideal" number of dimensions to be extracted. Such a number has to be fixed before the PRINCALS routine starts. In this study we tackled this problem by relying on a combination of several criteria: parsimony (few dimensions give a simpler description of data), accuracy (many dimensions fit data better), and interpretability (dimensions accounting for smaller proportions of variance tend to explain noise in data, rather than a systematic tendency). Accuracy is assessed through total and per-variable variance accounted for (VAF) indices, which provide the percentage of variance relating to the set of optimally scaled variables that is accounted for by the whole set of extracted dimensions (total VAF) and the single dimensions taken one at a time (per-variable VAF). The Cronbach alpha index [31] is also provided. It assesses the degree of internal consistency of (optimally scaled) variables and their relating indicator, to verify whether they jointly measure the same construct. This further supports the interpretations.

In this way, by applying PRINCALS to the set formed by the 4SQ scale and the perceived stress, control, and fatigue scales (ordinal data) [13-16], we obtained stress and control (numeric) indicators. As for lifestyle components, we obtained activity, smoking, and alcohol (numeric) indicators by applying PRINCALS to the whole set formed by the lifestyle qualitative variables (Table 1) and the quantitative estimation in METs of activity (Table 2).

Subsequently, with regard to step (3) (Figure 1), we employed the k-means clustering method [29] to construct subject typologies with respect to the following six components used as classification variables: stress and control indicators, lifestyle indicators (ie, activity, smoking, alcohol), and reported absenteeism (Table 2). The k-means clustering method is a nonhierarchical, iterative algorithm of classification that forms clusters (or groups) by minimizing the (squared) Euclidean distance between subjects and cluster centroids—that is, within-cluster vectors containing the means of variables. This is the same as forming clusters by minimizing the within-cluster deviance (ie, sum of squares).

This method has some well-known weaknesses: (1) final classification may depend on the order in which subjects appear in the data matrix, and (2) the number k of groups has to be fixed a priori. We addressed the problem of order dependency (problem 1) by employing a k-means cluster with *random starts*. A random start implies that the algorithm is initialized by choosing the k subjects, which have to represent the initial k clusters (so-called *seeds*), at random and without replacement. Then, the procedure iteratively attributes the remaining subjects to the nearest cluster on the basis of the squared Euclidean distance computed after updating the cluster centroids. Usually, it is advisable to adopt a great amount of random starts, and choose the partition that guarantees the minimum within-cluster deviance, so as to form groups that are as homogeneous as possible. Regarding problem (2), how to choose the ideal number of groups, we compute the Calinski-Harabasz (CH) statistic [32], which is given by the ratio of between-cluster variance to within-cluster variance. Both these variances derive from the division of the corresponding deviances by degrees of freedom equal to k in between-cluster variance and $N - k$ in

XSL•FO
RenderX

within-cluster variance, where *N* is the total number of subjects. Such a statistic is thus adjusted for the number of groups, and results derived from different classification schemes can be directly compared. The larger the statistic value, the greater the separation between groups, and the better the classification scheme pertaining to that specific partition in *k* groups.

An interesting feature of k-means clustering is its capacity to detect outliers—subjects with anomalous features with respect to the majority of data. If the algorithm is carried out as the number of groups increases, it may reveal small groups of isolated subjects that stably remain the same from a specific *k* onward. These small clusters can then be regarded as individuals or groups of outlying units, which can be removed and handled separately if proven to strongly affect results.

In this study, we performed k-means clustering (Figure 1) with 1000 random starts with the number of groups varying from $k = 2$ to $k = 15$ in two different phases. In the first phase, the algorithm was run with the specific goal of detecting potential outliers. To guarantee the same weight in the classification process, all six classification variables were standardized (*z* score) to have a mean of 0 (SD 1) before entering the clustering procedure. This shows the presence of six outliers (0.9% of the population), five falling in the same cluster plus one being isolated, which we therefore discarded in subsequent analyses. In the second phase, after removal of outliers, classification variables were standardized again. The algorithm was then performed as before on the remaining 677 participants. According to the CH statistic, seven is the optimal number of groups. Interpretation of clusters as subject typologies was carried out by means of boxplots of the within-clusters distribution of the classification variables. Typologies were labeled by the prevailing aspects that distinguished them from each other. In the absence of benchmarking and within the constraints of the present preliminary study, validation of groups was appraised with inferential procedures. Significance of differences between clusters was assessed with both parametric (univariate and multivariate ANOVA [MANOVA], and squared Mahalanobis distance test) and nonparametric (Kruskal-Wallis test) testing procedures [29,30], thus allowing for evaluation of the importance of all classification variables simultaneously, as well as one variable at a time. In particular, parametric procedures verified the hypotheses of equality of all cluster means for each single variable (ANOVA), equality of all cluster centroids (MANOVA), and equality of cluster centroids compared pairwise (squared Mahalanobis distance test). Kruskal-Wallis test, the nonparametric version of ANOVA, allows the equality of all cluster medians for each single variable to be checked.

Regarding step (4) (Figure 1), we first inferred the MeS status by the presence of at least three of the following ATP III criteria: (1) blood pressure ≥130/85 mmHg, (2) triglyceride levels ³150 mg/dL, (3) high-density lipoprotein cholesterol <40 mg/dL for males and <50 mg/dL for females, (4) fasting glycemia >100 mg/dL, and (5) waist circumference >102 cm for males and >88 cm for females. Participants with fewer than three factors were classed as preclinical, and those with no factors were classed as normal (Table 1).

Subsequently, we studied personal data (gender, work categories, and age group), illness, and MeS status with respect to subject typologies by computing within-typology and total percentages. Total percentages are computed for each variable on the whole set of participants without considering their aggregation in typologies. They can also be viewed as weighted arithmetic means of their corresponding within-typology percentages, where weights are given by the number of participants falling in the corresponding categories of the variables. Comparisons between total and within-typology percentages therefore reveal whether certain participants' characteristics tend to concentrate more highly (or more mildly) in specific groups than in the population. This analysis is mostly carried out in descriptive terms, by means of barcharts of within-typology and total percentage distributions computed for each variable. An overall chi-square test is first performed to verify whether subject typologies and those variables, considered one at a time, are statistically independent.

Subsequently, to learn more about the nature of associations between subject typologies and the various categories on personal data, illnesses, and MeS status, as revealed by chi-square tests, we computed standardized (or adjusted) Pearson residuals (APRs) [33]. For each 2-way contingency table obtained by cross-classifying subjects with respect to typologies and the above variables, APRs are given by the differences between corresponding observed and expected (ie, under the independence hypothesis) frequencies of subjects, which are then divided by their standard errors [33]. Given that APRs are asymptotically standard normal, inferences can be drawn, and significant single associations between typologies and categories of the above variables can be detected. Usually, an APR is considered "too great" to be consistent with a no-association hypothesis if it exceeds 2 or 3 in absolute value. Expressed in terms of (2-tailed) *P* values (taken from the standard normal), the two thresholds correspond, respectively, to $P = .046$ and $P = .003$. In this study, we introduced the value of 3.5 as well, which allows significant results to be detected at the level of $P < .001$.

Unless otherwise indicated, throughout this study the significance level was set at the .05 level. PRINCALS analysis, ANOVA, MANOVA, Kruskal-Wallis, Mann-Whitney, Kolmogorov-Smirnov, chi-square test, and APR were performed with SPSS version 18 (IBM Corporation, Somers, NY, USA); k-means clustering with random starts and graphics appearing next were carried out in the R environment, version 2.13.0 (R Foundation, Vienna, Austria); the squared Mahalanobis distance test was computed with SAS version 9.1 (SAS Institute, Cary, NC, USA).

Results

Below we present results of analyses following the four steps schematized in the Statistics section (Figure 1).

Summary Descriptive Statistics

Descriptive personal and lifestyle data of 683 participants are presented in Table 1.

XSL•FO
RenderX

Their modal age was 45–54 years, and gender was 72.5% (495/683) male and 27.5% (188/683) female. They were in large majority (629/683, 92.1%) white collar workers (only 3.7%, 25/683, blue collar workers and 4.2%, 29/683, managers) and were almost solely from one country (Italy). The majority of females (124/188, 66% females) were junior white collar workers, while a higher proportion of males than females occupied blue collar (24/495, 4.8% males vs 1/188, 0.5% females), senior white collar (197/495, 39.8% males, vs 59/188, 31.4% females), and management positions (25/495, 5.1% males, vs 4/188, 2.1% females).

Data show that age was unevenly distributed, although the distributions of males and females over age groups are similar. More than 70% (481/683) participants were nonsmokers, in accord with the smoking ban of the company. Almost 40% (270/683) declared an absence of physical activity; the majority (374/683, 54.8%) did not drink or remained within 1 to 2 glasses of wine or beer per week. Almost 80% (539/683) did not drink any alcohol, especially females (178/188, 94.7% females).

About 12% (82/683) of respondents said they had functional disturbances, and almost 25% (169/683) reported some form of chronic disease; the majority, however (432/683, 63%), said they had no active diseases. There were no apparent differences between males and females.

Almost 49% (331/683) of participants (preclinical) reported one or two risk factors for MeS, and 23% (157/683) had MeS status. Most females (140/188, 74.5%) declared no risk factors, while the majority of males appeared preclinical (285/495, 57.6%) or within MeS (155/495, 31.3%). Chi-square test showed significant differences between males and females as regards work categories, alcohol habit (wine or beer, and alcohol in glasses per week), and MeS status (see Table 1).

Table 2 summarizes statistics on reported biochemical and anthropometric data, corroborating that, as expected from a working population prevalently comprising middle-aged workers, mean population values are reasonably within or near normal limits. Estimated weekly METs of activity were quite low owing to substantial lack of moderate or vigorous activity in a large fraction (about 25%, first quartile) of the study population. Regarding reported absenteeism, yearly lost days were overall very low (mean 5.87 days/year) with a single notable exception (300 days). Average indices of perceived bodily symptoms, stress, fatigue, and control were all within the range that can be observed in normal individuals in our laboratory.

Parametric and nonparametric testing procedures (see Table 2) agree on supporting significant differences between males and females in all reported biochemistry (except total cholesterol), as well as blood pressure and anthropometrics. Vigorous and total activity, reported absenteeism, and 4SQ, perceived stress, and fatigue scales were also significantly different.

Setting Up of Stress, Control, and Lifestyle Indicators

To account for the multivariate nature of domains under study, we applied the PRINCALS method to the two sets formed by (1) the four stress and control perception scales, and (2) lifestyle variables, thus obtaining, respectively (1) two dimensions, which can be interpreted as indicators of stress and control (total VAF: 84.2% of variance of the four optimally transformed scales of self-reported stress and control; total Cronbach alpha = .937) and (2) three dimensions interpreting lifestyles, which can be regarded as indicators of activity, smoking, and alcohol habit (total VAF: 70.1% of all optimally transformed lifestyle variables; total Cronbach alpha = .947).

These interpretations of the indicators derive from the analysis of component loadings. In particular, component loadings for the stress indicator are .885 with the 4SQ scale, .885 with the perceived fatigue scale, .870 with the perceived stress scale, and .310 with the perceived control scale. This denotes a strong linear relationship of the stress indicator with the first three scales, while the link with the control domain turns out to be negligible. Notably, the stress indicator accounts for 60.5% of the total variance. Moreover, Cronbach alpha computed for the (optimally transformed) 4SQ, perceived stress, and fatigue scales is equal to .934 if the stress indicator is involved in computations, but .865 if the stress indicator is disregarded. The three scales and their indicators are therefore characterized by a high internal consistency.

On the other hand, the component loadings for the control indicator are equal to –.033 with the 4SQ scale, –.172 with the perceived fatigue scale, –.129 with the perceived stress scale, and .949 with the perceived control scale. This suggests interpreting the second dimension in terms of the control indicator, which accounts for 23.7% of the total variance.

Interpretation of lifestyle dimensions as indicators is based on the component loadings reported in Table 3. It is apparent that dimension 1 is highly positively correlated with activity variables (VAF: 32.4%), dimension 2 with smoking variables (VAF: 21. 9%), and dimension 3 with alcohol variables (VAF: 15.8%). Moreover, Cronbach alpha values computed for each subset of the lifestyle variables (denoted with a, b, and c in Table 3) and the corresponding indicator are high (Table 3, next to last row), thus proving high internal consistency in all cases.

XSL•FO
RenderX

Table 3. Component loadings for lifestyle indicators

Lifestyle (quantified) variables	Lifestyle indicators		
	Dimension 1[a]	Dimension 2[b]	Dimension 3[c]
Smoking habit	–.067	.961[b]	–.200
Intend to quit	–.079	.958[b]	–.205
Estimated metabolic equivalents of activity			
Walking	.571[a]	.102	.078
Moderate activity	.710[a]	–.029	–.085
Vigorous activity	.773[a]	.043	–.040
Total activity	.975[a]	.057	–.032
Frequency of structured physical activity	.716[a]	–.032	–.112
Wine or beer (glasses/week)	.094	.202	.816[c]
Alcohol (glasses/week)	.065	.265	.802[c]
Variance accounted for	32.4%	21.9%	15.8%
Cronbach alpha			
With dimension	.882	.983	.868
Without dimension	.808	.967	.615

[a,b,c] Component loadings with absolute value >0.4. a: dimension 1 = activity indicator, b: dimension 2 = smoking indicator, c: dimension 3 = alcohol indicator.

Construction of Subject Typologies

The k-means clustering method with 1000 random starts was employed to form clusters of participants according to their scores on the indicators of stress, control, activity, smoking, and alcohol, and their reported absenteeism, all used as classification variables in standardized form. Table 4 reports their minimum and maximum values and quartiles. By comparing the maximum value with the third quartile, it is apparent that outlying participants (in a univariate sense) are present in the dataset. For instance, regarding reported absenteeism, 75% of participants had *z* scores equal at most to 0.0091, while the maximum value is 19.88. This clearly denotes the presence of at least one outlier. A similar argument can be advanced for the alcohol indicator and, though less apparently, also for the stress, activity, and smoking indicators.

Table 4. Descriptive data of classification variables (N = 683 participants)

Indicators (*z* score)	Minimum	Maximum	1st quartile	2nd quartile	3rd quartile
Stress	–1.2404	3.3452	–0.7531	–0.3071	0.5719
Control	–2.1358	1.4964	–1.0103	0.0668	1.0068
Activity	–1.3373	4.6326	–0.7692	–0.2112	0.5596
Smoking	–0.6954	4.5797	–0.5658	–0.4533	–0.1202
Alcohol	–1.7381	10.6010	–0.5668	–0.2588	0.4445
Reported absenteeism	–0.3964	19.8797	–0.3964	–0.1937	0.0091

Moreover, standardized classification variables (Table 4) proved overall to have very low linear correlations (table omitted for simplicity), so no multicollinearity problem arose. This excludes the drawback of more highly correlated variables exerting higher weights on the classification procedure. By construction the three indicators of lifestyles are uncorrelated with each other, as well as the two indicators of stress and control. With regard to cross-comparisons between different sets of variables, the two highest observed correlations, in absolute value, concern reported absenteeism and stress indicator (0.24), and activity indicator and stress indicator (–0.17). The other correlations, being close to zero, are negligible.

Figure 2 shows values of the CH statistic obtained by repeatedly applying the k-means algorithm with the number of groups increasing from $k = 2$ to $k = 15$. A first run (Figure 2, circles) of this procedure revealed that the partition guaranteeing the maximal separation of groups is that formed by seven clusters (CH = 154.93). Two among these, however, were very small groups, which involved 6 participants in all (6/683, 0.9%), 5 falling in the same cluster plus 1 being isolated. In particular,

XSL•FO
RenderX

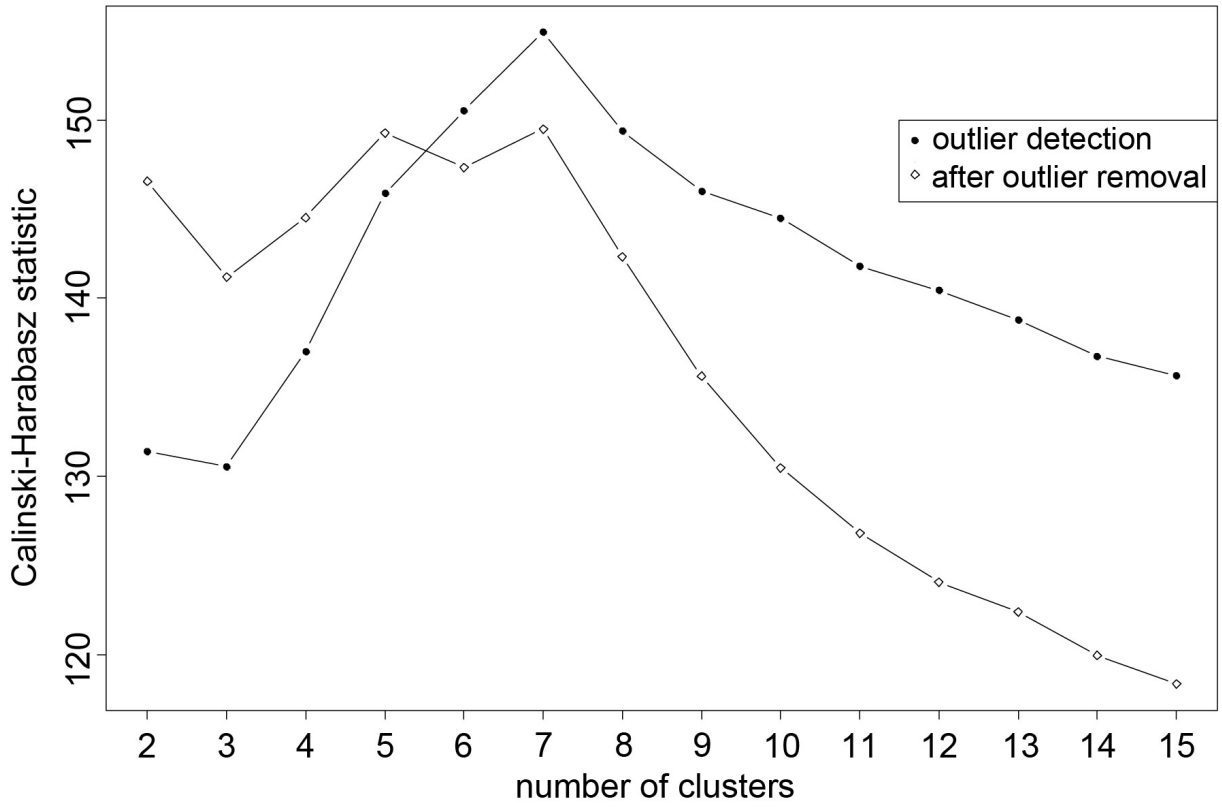
this latter participant (male) was featured by a very high absenteeism score (300 lost working days, which corresponds to a *z* score of 19.88), a very high stress score (3.27), and a very low activity score (–1.14). Moreover, he was in the 45–54 year age group, was a junior white collar worker, and reported having an organic illness. The other 5 participants (4 males, 1 female) were characterized by the highest scores on alcohol and smoking indicators. In addition, they had the highest scores for stress and absenteeism; 4 of these participants reported a functional illness and the other an organic illness; 3 were in the 45–54 year age group and the other 2 were more than 54 years old; 1 was a blue collar worker, 2 were junior white collar workers, and 2 were senior white collar workers.

As these two small groups constantly kept the same structure while the number of groups increased, we consider them to be formed by outliers, and accordingly we removed them from subsequent analyses. We then restandardized the classification variables on the remaining 677 participants. After outlier removal, correlations between classification variables were practically unchanged. A reduced correlation of reported absenteeism and stress indicator (0.19) is the unique appreciable variation.

The k-means clustering procedure with 1000 random starts, performed under these new circumstances, produced a new set of values on the CH statistic, which we computed, once again, as the number of groups varied from $k = 2$ to $k = 15$ (Figure 2, diamonds). This figure clearly illustrates that the CH statistic assumes the highest values at five clusters (CH = 149.28) and seven clusters (CH = 149.49). After careful consideration we opted for seven clusters, as compared with five (data not shown for simplicity), because this classification better represents the main differences, as well as similarities, between participants.

The first three columns of Table 5 show information about cluster sizes. Numeric cluster labels (first column) are automatically assigned by the clustering procedure, without any specific meaning. The biggest cluster (cluster 5) contains 194 participants (194/677, 28.7%), while the smallest (cluster 7) consists of 20 (20/677, 3%).

Figure 2. Calinski-Harabasz statistic in k-means clustering with 1000 random starts. First phase (line with circles): detection of outliers; second phase, after outlier removal (line with diamonds): search for optimal number of clusters.



XSL•FO
RenderX

Table 5. Cluster size and description of subject typologies

Cluster	Count	Percentage	Description of typology	Typology label
1	90	13.3	Highest levels of alcohol habit; mostly nonabsentees, nonsmokers; control indicator highly variable	Alcohol
2	98	14.5	Highest levels of smoking habit; mostly non-physically active, nondrinkers, nonabsentees; control indicator highly variable	Smoking
3	88	13.0	Highest levels of stress; mostly lower levels of control, non-physically active, nonsmokers	High stress
4	57	8.4	Highest levels of physical activity; mostly lower levels of stress, nonsmokers, nondrinkers, nonabsentees	Physical activity
5	194	28.7	Highest levels of control; mostly lower levels of stress, nonsmokers, nonabsentees	High control
6	130	19.2	Lowest levels of stress and control; mostly nonsmokers, nonabsentees	Low stress and control
7	20	3.0	Highest levels of absenteeism; mostly nonsmokers, nondrinkers; stress and control indicators highly variable	Absenteeism
Total	677	100.0		

Interpretation of clusters in terms of subject typologies is derived from the analysis of boxplots of within-cluster distributions (Figure 3) of the standardized classification variables. In particular, remarkable associations with specific groups can be observed, such as cluster 3, in which high stress might be associated with low activity, and conversely, a high level of physical activity might be associated with low level of stress in cluster 4. However, similar patterns may not be apparent across the entire survey population, thus suggesting that the relationship between stress and physical activity can assume various forms, especially if considered in the presence of other participants' characteristics—for example, work category, age, or presence of illnesses.

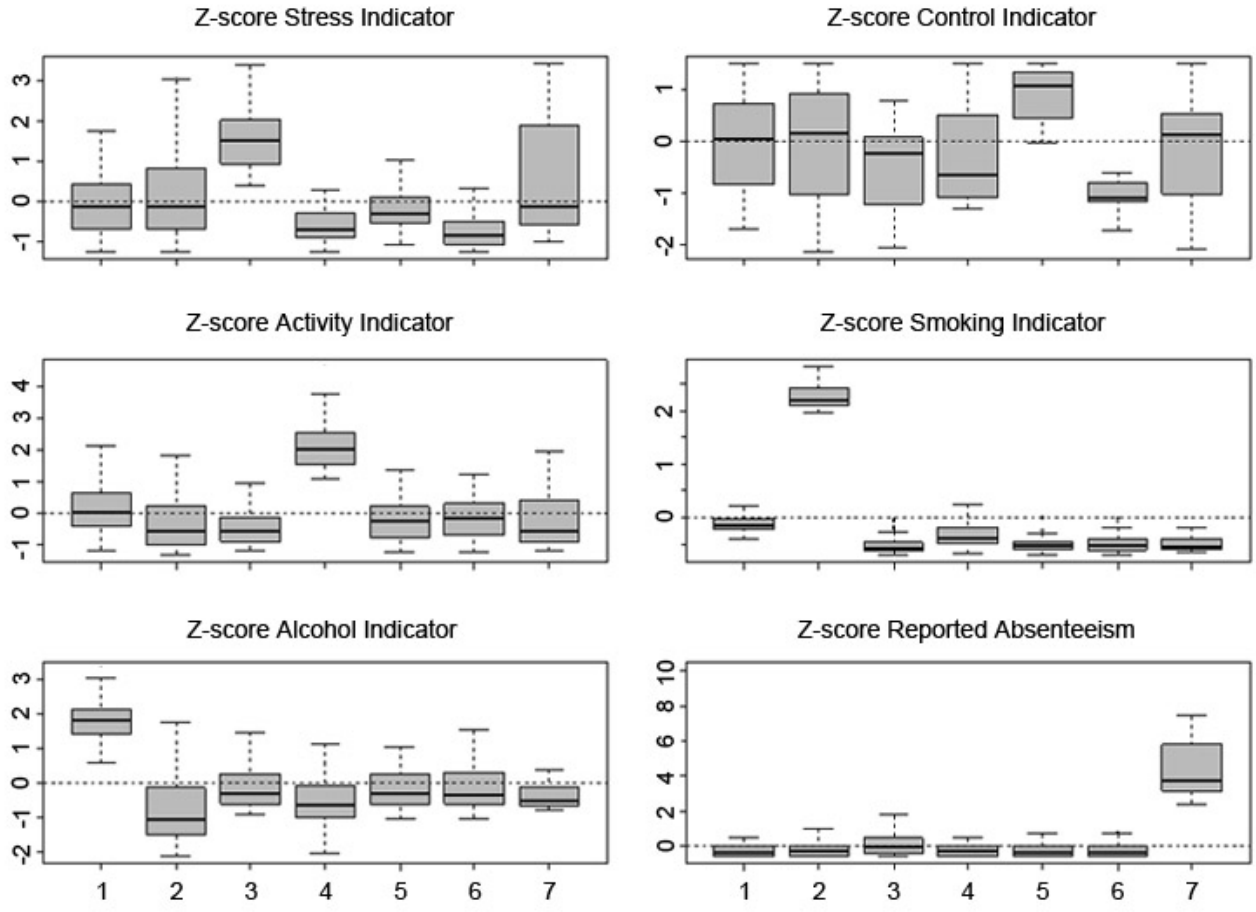
Differences in absenteeism across clusters are limited, with the exception of cluster 7, which contains all the participants with the greatest number of lost working days. As for stress, nearly 50% of participants in the cluster with the highest levels of stress (cluster 3) reported some absenteeism. The other clusters, including the one with the highest reported activity (cluster 4), show very low levels of absenteeism.

Finally, hypothesis testing procedures, carried out with both parametric and nonparametric methods, empirically supported significant differences between the seven clusters. In detail, the MANOVA procedure led us to reject the null hypothesis of equal cluster centroids with all six classification variables considered simultaneously (Wilks' lambda: $P < .001$; Hotelling-Lawley trace: $P < .001$). Also, the squared Mahalanobis distance test led us to reject the hypothesis of equality between cluster centroids in all 21 possible comparisons between couples of groups ($P < .001$). Finally, for each variable, univariate ANOVA led us to reject ($P < .001$) the hypothesis of equal cluster means, and Kruskal-Wallis test led us to reject the hypothesis of equal cluster medians ($P < .001$).

The last two columns in Table 5 summarize the findings of cluster characterization in terms of subject typologies. They contain a summary description of specific characteristics of these clusters and the labels of the typologies they represent.

XSL•FO
RenderX

Figure 3. Boxplots of within-cluster distributions of standardized classification variables (x-axis: numeric cluster labels as given in Table 5).



Relationships Between Subject Typologies and Personal Data, Illness, and Metabolic Syndrome

The final step of our study consisted of assessing potential relationships between subject typologies and each of personal data, illness status, and MeS. This analysis was mostly carried out in descriptive terms through computation and then comparison of within-typology and total percentage distributions.

Figure 4 presents panels of barcharts of within-typology and total percentages, the latter of which were computed for each variable on the set of 677 participants remaining after outlier deletion, without considering classification. Several worthwhile aspects are detailed in Multimedia Appendix 1, and schematically indicated below. Table 6 summarizes the main features that distinguish the subject typologies in terms of the major (or minor) concentration of participants with certain characteristics as compared with the survey population.

Chi-square test empirically supported significant associations between subject typologies and personal data, illness status, and MeS condition. Specifically, overall associations between subject typologies and either gender, work categories, illness, or MeS status were all significant at the .001 level; association with age group was significant at the .05 level (P = .02). The more thorough analysis subsequently carried out with APR (Figure 4, boxed symbols – and +) highlighted that specific associations between single typologies and categories of variables were stronger than expected under statistical independence. In particular, the overall significant relationship between gender and subject typologies appeared to substantially arise from the significant associations involving the alcohol cluster (more males than expected) and the high stress cluster (more females than expected). Moreover, the alcohol cluster included fewer participants without risk factors for MeS and more with preclinical MeS than expected. The high stress cluster turned out to be significantly positively associated with workers aged 35–44 years, junior white collar workers, respondents reporting functional or organic illnesses, and those without MeS. Conversely, the high stress cluster was significantly negatively associated with senior white collar workers. The physical activity cluster was significantly positively associated with participants without illness or without risk factor for MeS. It was significantly negatively associated with senior white collar workers and participants with MeS. The absenteeism cluster proved to be significantly positively associated with older participants (>54 years), blue and junior white collar workers, and participants with organic illnesses. Conversely, it was significantly negatively associated with senior white collar workers and those without illnesses. Finally, the low stress and control cluster was significantly positively associated with participants without illnesses and negatively with those reporting functional illnesses.

XSL•FO
RenderX

Figure 4. Barcharts of within-typology percentage distributions of personal data (gender, work categories, age group), illness status, and metabolic syndrome (MeS). Details regarding statistical symbols and significance are reported in Multimedia Appendix 1. (Phys = physical, Strs&Ctrl = stress and control; Tot. perc. = total percentage).

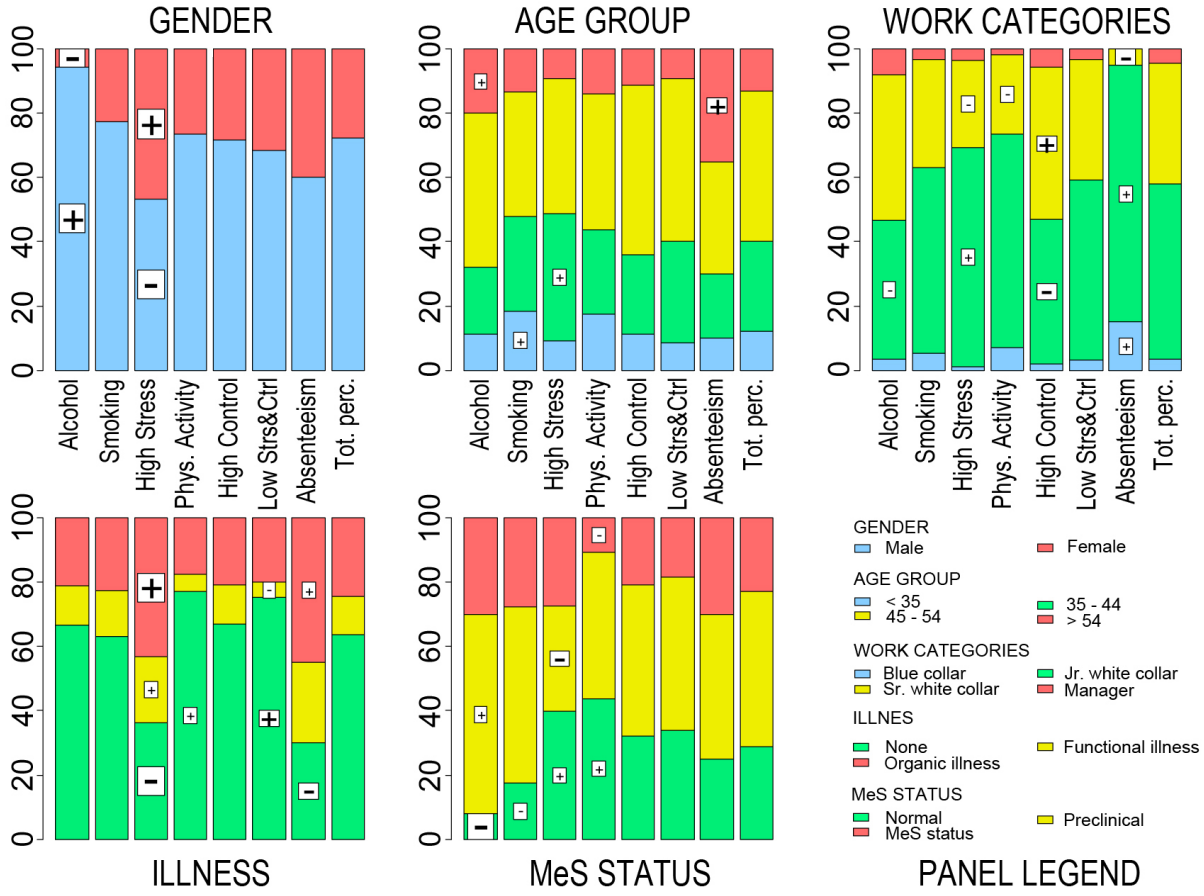


Table 6. Composition of subject typologies with respect to personal data, illnesses, and metabolic syndrome condition, and prevailing characteristics with respect to the distribution (more or less) of the considered characteristic in the survey population (descriptive analysis)

Typology	Label	Composition
1	Alcohol	More males, senior white collar workers, managers, with preclinical MeS[a] or MeS, >54 years old
2	Smoking	More <35 years old
3	High stress	More females, junior white collar workers, 35–44 years old, with functional or organic illnesses, without MeS; less blue collar workers, >54 years old, preclinical MeS
4	Physical activity	More blue collar and junior white collar workers, <35 years old, healthy, without MeS; less with functional or organic illnesses, with MeS
5	High control	More senior white collar workers and managers, 45–54 years old
6	Low stress and control	More healthy, 35–54 years old; less <35 years old, with functional or organic illnesses
7	Absenteeism	More females, blue collar and junior white collar workers, >54 years old, with functional or organic illnesses, with MeS; less 35–54 years old; no managers

[a] Metabolic syndrome.

Discussion

This study shows the feasibility of assessing health profile, lifestyles, and work habits using an ad hoc self-administered questionnaire via an Intranet application of a large company. The basic ingredients of such an assessment consist in constructing numeric indicators, whichever is the nature of the available information (qualitative and/or quantitative), forming subject typologies from how the indicators combine, and investigating the composition of subject typologies with respect to external variables (ie, variables not involved as classification variables). The main potential of this approach is that it does not require imposing any functional form to relationships between variables, given that, if present, these relationships are learned directly from the data.

XSL•FO
RenderX

Statistics

A strength of our study is that the use of advanced multivariate statistical methods (ie, PRINCALS [28] and k-means clustering [29] methods) allows the derivation from self-reported data of a series of numeric indicators describing several unobservable variables [34], such as perceived stress and control, and lifestyle domains, as well as the construction of subject typologies and examination of their possible relationship with personal data, illness, and MeS status.

Regarding the specific potentials, unlike statistical–probabilistic modeling, our data-driven approach did not require us to specify functional forms for the relationships between variables, which would have been inappropriate in this context due to the exploratory purposes of the study. In addition, this approach allowed us to get around possible limitations inherent in the available information, especially participants' self-selection. Being exploratory, these methods do not aspire to generalize findings to sets of subjects not expressly involved in the survey. Their descriptive range is confined within "what is actually observed," so it does not really matter whether subjects are self-selected or not. Collected data are treated as the unique reference population, whose characteristics are then synthesized, described, and interpreted.

However, in the study we did not completely discard statistical testing procedures. Undoubtedly, the nonidentifiability of the target population implies that test results should be interpreted with some caution, since it is not clear to which population the drawn inferences have to be referred. This is a critical point shared by most surveys, especially Web surveys. Even if the reference population should be a priori well defined (such as in our case, where the reference population is given by the 24,000 workers of the Italian branch of the company), a large number of nonrespondents, typically occurring in these kinds of surveys, would make the set of respondents not representative of the entire population. Nonetheless, statistical tests may help reveal crucial relationships deserving more careful consideration, as well as give rise to new research conjectures which ad hoc studies should address in future investigations. This is the main reason why we have performed inferential analyses as well.

As we claimed, we have performed both parametric and nonparametric testing procedures to take advantage of their specific potentials [29,30]. As it is known in statistical literature, parametric tests may lead to unbiased conclusions, either if data are far from being normally distributed or if any other basic assumption fails to hold, such as the requirement of homogeneity of within-group variances in univariate ANOVA. In most situations the distribution of quantitative data can be rendered approximately normal, or within-group variances can be made homogeneous, by appropriate transformations—for example, by computing the logarithm of values of each variable. However, this procedure may complicate interpretations of results, since these latter have to be referred to transformed, instead of original, data. Conversely, nonparametric procedures, being distribution-free, are not sensitive to departures from normality. They are recognized, however, to be generally less powerful than parametric procedures, in the sense that, for a fixed nominal significance level, nonparametric tests lead to acceptance of the alternative hypothesis when it is true with a lower probability than a parametric test. For these reasons, we have decided to rely on both parametric and nonparametric methods, and then we have considered a test result as "sufficiently revealing" if borne out as significant by both procedures.

Assessing Stress at Work

Stress is a ubiquitous component of everyday activities, affecting both work and private life. Interest in its assessment has recently increased in view of the tight relationship with a number of negative consequences, either in the subjective domain, such as perceived quality of work and absenteeism, or in the clinical domain, impairing risk profile particularly in the cardiometabolic area [8,35,36]. The majority of stress tools provide metrics that are based on self-reports with standardized questionnaires that are intrinsically prone to bias. These tools are being modified to better focus on stress at work, also in view of the recent policies that, in many countries, mandate stress assessment at work. Usually this approach focuses more on organizational aspects (following motivational models such as the demand-reward [37] or job strain [38]) than on individual physiopathological consequences, such as the increase in sympathetic drive or in hormonal burden (eg, increased cortisol secretion [12]). This selective window might only slightly impair the determination of organizational stress at companywide levels, but may be suboptimal for gathering information useful to designing and planning individualized strategies to tame stress and its health consequences, such as hypertension or worsening metabolic risk. These tools, although simultaneously addressing various aspects of people's behavior, usually do not employ multivariate statistical techniques, jointly combining ordinal and quantitative data. In previous studies [13-16] we combined information from self-reports focusing on symptom profile and simple indicators of perception of stress, fatigue, and control, with physiopathological data consisting of simple hemodynamics (heart rate and blood pressure) and autonomic indices from heart rate variability. This approach proved valuable to appraise the elevated stress attending a companywide reorganization and to demonstrate the effectiveness of lifestyle strategies to manage stress [13]. In a different study regarding stress management in a clinical setting, in order to curtail the error bias resulting from the inherent imprecision of the subjective measures, we employed a modified approach based on the computation of hidden factors, improving the accuracy of metrics describing a combined stress dimension from multiple indicators [16]. The approach presented in the present study, based on an IT instrument, and on multivariate statistics might prove more robust as a tool to assist individual adherence to self-managed programs for lifestyle improvement and risk reduction, as mandated by several recent guidelines in the hypertension [21] or cardiology area [19].

Stress and Lifestyle

Because of the large error expected to potentially affect single variables pertaining to ill-defined concepts such as stress, to make allowance for the possible bias of the technique employed (self-reports and unsupervised questionnaire), and to address the multivariate nature of the domains under study (eg, 4SQ, and perceived stress and fatigue scales are expected to be linked

XSL•FO
RenderX

together by inextricable interrelationships), we set up indicators of stress, control, and lifestyles, considering in particular activity, smoking, and alcohol habit. In this way it was possible to enhance the information extracted from the data set by synthesizing them in an optimal sense and limiting potentially redundant semantic overlaps. The use of the CH statistic allowed us to select the optimum number of subject typologies that emerged from the analysis and that revolved around few key indicators: alcohol and smoking habits, stress, activity, control, and absenteeism. This approach might thus extend the stress model, which we used for several years in multiple studies involving volunteers or patients [14-16], as well as workers [13], and which provided consistent results, to Web-based self-administered applications.

The present approach evaluates the relationship between typologies and personal data, allowing exploration of key aspects of health promotion and prevention strategies in a normal working population.

For example, stress has been hypothesized as a component or modulator in MeS [36]. In this study, high stress was more prevalent in females, and was observed slightly more in younger participants and junior white collar workers. Thus, stress in this specific population might promote unhealthy behavior, not so much through smoking and poor nutrition, but through inactivity, particularly in the female population [8,39,40]. High stress is also observed more easily in respondents reporting the presence of illness.

Conversely, in the MeS profile, stress unexpectedly showed an elevated percentage of normal. We might interpret this finding to indicate that in the initial phases stress may be perceived subjectively by younger workers, but metabolic implications might require the influence of additional factors over time, such as inactivity favoring the occurrence of obesity, but which cannot be observed in this exploratory investigation. Alcohol abuse could instead play a significant role in facilitating the preclinical condition of MeS. As a final consideration regarding potential validity of the present data, the MeS prevalence in the examined population is similar to that reported for general populations (eg, in the United States [41]).

Limitations

Because some investigators cast doubts on the validity of self-reported, as compared with non-self-reported, data [11], a few comments seem warranted, particularly considering the specific condition of Web-based applications designed as a part of personalized preventive strategies in the workplace [13]. First, let us consider that every kind of data (Y), either non-self-reported or self-reported, can be conceptualized [11] as the sum of the following factors: “true” data (TD), plus systematic bias (SB) and random error (RE), according to the formula $Y= \lambda_{TD}TD + \lambda_{SB}SB + RE$, where λ denotes factor loadings.

It should also be noted that even non-self-reported data are not equivalent to error-free data: even simple transcription from paper forms may lead to about a 3% to 26% error rate [42]. Regarding bias, we should consider that it may differ according to the specific context or variable involved—for example, behavioral multicomponent constructs (exercise, stress, etc) may counterintuitively be more accurately represented by self-reports because of the lack of (potentially greater interpretative) bias introduced by a third party (physician, nurse, or technician) [11] and because subjects are likely to better interpret questions about their own behavioral conditions.

Regarding biochemical data, blood pressure, or anthropometrics, we feel confident that only a relative small bias could characterize the self-reported data of this Web-based study, considering that participants were digitally competent and highly motivated to follow the instructions accurately because the usefulness of the final report was contingent on the quality of input data. In this sense, in certain cases when self-reports are the sole source of information, they have been considered “invaluable”, as in the case of the National Health Interview Survey [43]. Moreover, it has been said that Internet-based studies, with a particular focus on self-selection, are of at least as good a quality as those provided by traditional paper-and-pencil methods [44]. For these reasons we avoided putting too much emphasis on single biochemical data, but we combined them with anthropometrics and blood pressure to estimate, as a synthetic descriptor, the MeS status, which, also in our hands [22,27], proved very useful in exploratory population studies.

In short, we are confident that possible errors, if present, are unsystematic, in the sense that they are not in the same direction or with the same magnitude, or else we should suppose that respondents agreed on hindering the survey. For the goal of this study, we therefore considered the constructed indicators and the derived subject typologies to be reliable.

Tools to assess general health and cardiovascular risk, based on a multivariate algorithm, are widely used [45,46], and their main goal is to guesstimate the probability of developing an event in a given time window (usually 5 or 10 years). Conversely, the declared goal of this Web-based questionnaire was to indicate to participants the presence or absence of areas of potential cardiometabolic risk, which could merit a focused improvement, even if the computed global risk was low (as is easily the case for young participants with just one or two usual risk factors). The same applies to those with initial levels of established risk factors, such as prehypertension or nonoptimal lipid profile, frequently combined as forerunners of MeS [46-48]. In this case the use of lifestyle-based interventions could be particularly beneficial, with very low cost and no unwanted effects [19]. The present Internet instrument, although probably suboptimal in providing “hard” health information, might instead be very useful in assisting “soft” lifestyle changes by evoking in participants the awareness and motivation [49] needed to obtain a long-term change in behavior, thus adding an educational flavor to the effectiveness of professional help. Finally, although providing intervention scores to respondents might facilitate compliance, it could also bias the overall data base. Very low occurrence of duplicate responses, however, suggests that this bias was probably minimal.

Additional caution should finally be used in evaluating quantitatively presented data, as other factors, such as the

XSL•FO
RenderX

nonsmoking policy of the company, might bias findings, particularly in regard to their external validity.

Implications for Prevention

In the field of early primary prevention the active role of the individual and the coaching role of the employer have been amply discussed [50]. While the current standard of medical practice for acute conditions relies heavily on institutional resources, prevention must face the challenge of long-term, patient-driven behavioral modifications, based on an agreement on lifestyle determinants. In this model, digital techniques are useful to obtain a streamlined flow of information between patients (or rather people) and the various stakeholders, having employers in the front line of investment. The present investigation showing the feasibility of assessing subject typologies and their relationship with personal characteristics at a workplace with a simple Intranet application might suggest that the time is ripe to test large-scale applications of information and computer technology for better detection and treatment of cardiometabolic risk at the population level, as a complementary benefit offered at the employer's cost. Studies such as the present one might provide additional momentum to further IT applications as tools for health promotion in the workplace.

Acknowledgments

We would like to acknowledge the expert secretarial help of Giovanna Maccio'. Supported by ASI contract DCMC, PRIN 2007.

Conflicts of Interest

None declared

Multimedia Appendix 1

More detailed description of data in Figure 4.

[PDF File (Adobe PDF File), 52KB - jmir_v13i4e88_app1.pdf]

References

1. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 2004;364(9438):937-952. [doi: 10.1016/S0140-6736(04)17018-9] [Medline: 15364185]
2. Rosengren A, Hawken S, Ounpuu S, Sliwa K, Zubaid M, Almahmeed WA, INTERHEART investigators. Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. Lancet 2004;364(9438):953-962. [doi: 10.1016/S0140-6736(04)17019-0] [Medline: 15364186]
3. Daubenmier JJ, Weidner G, Sumner MD, Mendell N, Merritt-Worden T, Studley J, et al. The contribution of changes in diet, exercise, and stress management to changes in coronary risk in women and men in the multisite cardiac lifestyle intervention program. Ann Behav Med 2007 Feb;33(1):57-68. [doi: 10.1207/s15324796abm3301_7] [Medline: 17291171]
4. Vartiainen E, Laatikainen T, Peltonen M, Juolevi A, Männistö S, Sundvall J, et al. Thirty-five-year trends in cardiovascular risk factors in Finland. Int J Epidemiol 2010 Apr;39(2):504-518 [FREE Full text] [doi: 10.1093/ije/dyp330] [Medline: 19959603]
5. Kotseva K, Wood D, De Backer G, De Bacquer D, Pyörälä K, Keil U, EUROASPIRE Study Group. Cardiovascular prevention guidelines in daily practice: a comparison of EUROASPIRE I, II, and III surveys in eight European countries. Lancet 2009 Mar 14;373(9667):929-940. [doi: 10.1016/S0140-6736(09)60330-5] [Medline: 19286092]
6. Jones DW, Peterson ED, Bonow RO, Gibbons RJ, Franklin BA, Sacco RL, American Heart Association. Partnering to reduce risks and improve cardiovascular outcomes: American Heart Association initiatives in action for consumers and patients. Circulation 2009 Jan 20;119(2):340-350 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.191328] [Medline: 19124667]
7. Wilcox S, Dowda M, Leviton LC, Bartlett-Prescott J, Bazzarre T, Campbell-Voytal K, et al. Active for life: final results from the translation of two physical activity programs. Am J Prev Med 2008 Oct;35(4):340-351. [doi: 10.1016/j.amepre.2008.07.001] [Medline: 18779028]
8. Dimsdale JE. Psychological stress and cardiovascular disease. J Am Coll Cardiol 2008 Apr 1;51(13):1237-1246. [doi: 10.1016/j.jacc.2007.12.024] [Medline: 18371552]
9. Medicine ACOPSOI, Thompson WP, Gordon N, Pescatello LS. ACSM's Guidelines for Exercise Testing and Prescription. 7th edition. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2010.
10. Lucini D, Pagani M. Exercise: Should it matter to internal medicine? Eur J Intern Med 2011 Aug;22(4):363-370. [doi: 10.1016/j.ejim.2011.02.022] [Medline: 21767753]
11. Chan D. So why ask me? Are self-report data really that bad? In: Lance CE, Vandenberg RJ, editors. Lance CE, Vandenberg RJ, editors. Statistical and Methodological Myths and Urban Legends. New York, NY: Routledge; 2009.

XSL•FO
RenderX

12. McEwen BS. Protective and damaging effects of stress mediators. N Engl J Med 1998 Jan 15;338(3):171-179. [doi: 10.1056/NEJM199801153380307] [Medline: 9428819]
13. Lucini D, Riva S, Pizzinelli P, Pagani M. Stress management at the worksite: reversal of symptoms profile and cardiovascular dysregulation. Hypertension 2007 Feb;49(2):291-297 [FREE Full text] [doi: 10.1161/01.HYP.0000255034.42285.58] [Medline: 17210835]
14. Lucini D, Norbiato G, Clerici M, Pagani M. Hemodynamic and autonomic adjustments to real life stress conditions in humans. Hypertension 2002 Jan;39(1):184-188 [FREE Full text] [Medline: 11799100]
15. Lucini D, Di Fede G, Parati G, Pagani M. Impact of chronic psychosocial stress on autonomic cardiovascular regulation in otherwise healthy subjects. Hypertension 2005 Nov;46(5):1201-1206 [FREE Full text] [doi: 10.1161/01.HYP.0000185147.32385.4b] [Medline: 16203875]
16. Lucini D, Malacarne M, Solaro N, Busin S, Pagani M. Complementary medicine for the management of chronic stress: superiority of active versus passive techniques. J Hypertens 2009 Dec;27(12):2421-2428. [doi: 10.1097/HJH.0b013e3283312c24] [Medline: 19887956]
17. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, American College of Sports Medicine, American Heart Association. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Circulation 2007 Aug 28;116(9):1081-1093 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.185649] [Medline: 17671237]
18. 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. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
19. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, American Heart Association Strategic Planning Task Force and Statistics Committee. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. Circulation 2010 Feb 2;121(4):586-613 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.109.192703] [Medline: 20089546]
20. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
21. Mancia G, De Backer G, Dominiczak A, Cifkova R, Fagard R, Germano G, ESH-ESC Task Force on the Management of Arterial Hypertension. 2007 ESH-ESC Practice Guidelines for the Management of Arterial Hypertension: ESH-ESC Task Force on the Management of Arterial Hypertension. J Hypertens 2007 Sep;25(9):1751-1762. [doi: 10.1097/HJH.0b013e3282f0580f] [Medline: 17762635]
22. Malacarne M, Gobbi G, Pizzinelli P, Lesma A, Castelli A, Lucini D, et al. A point-to-point simple telehealth application for cardiovascular prevention: the ESINO LARIO experience. Cardiovascular prevention at point of care. Telemed J E Health 2009 Jan;15(1):80-86. [doi: 10.1089/tmj.2008.0066] [Medline: 19199851]
23. Ritterband LM, Tate DF. The science of internet interventions. Introduction. Ann Behav Med 2009 Aug;38(1):1-3. [doi: 10.1007/s12160-009-9132-5] [Medline: 19816750]
24. Neville LM, O'Hara B, Milat A. Computer-tailored physical activity behavior change interventions targeting adults: a systematic review. Int J Behav Nutr Phys Act 2009;6:30 [FREE Full text] [doi: 10.1186/1479-5868-6-30] [Medline: 19490649]
25. Verheijden M, Bakx JC, Akkermans R, van den Hoogen H, Godwin NM, Rosser W, et al. Web-based targeted nutrition counselling and social support for patients at increased cardiovascular risk in general practice: randomized controlled trial. J Med Internet Res 2004 Dec 16;6(4):e44 [FREE Full text] [doi: 10.2196/jmir.6.4.e44] [Medline: 15631968]
26. Bennett GG, Herring SJ, Puleo E, Stein EK, Emmons KM, Gillman MW. Web-based weight loss in primary care: a randomized controlled trial. Obesity (Silver Spring) 2010 Feb;18(2):308-313 [FREE Full text] [doi: 10.1038/oby.2009.242] [Medline: 19696764]
27. Grundy SM, Brewer B, Cleeman JI, Smith SC, Lenfant C, National Heart, Lung, and Blood Institute, American Heart Association. Definition of metabolic syndrome: report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. Arterioscler Thromb Vasc Biol 2004 Feb;24(2):e13-e18 [FREE Full text] [doi: 10.1161/01.ATV.0000111245.75752.C6] [Medline: 14766739]
28. Gifi A. Nonlinear Multivariate Analysis. New York, NY: Wiley; 1990.
29. Dillon WR, Goldstein M. Multivariate Analysis: Methods and Applications. New York, NY: Wiley; 1984.
30. Hollander M, Wolfe DA. Nonparametric Statistical Methods. New York, NY: Wiley; 1999.
31. Cronbach LJ. Coefficient Alpha and the internal structure of tests. Psychometrika 1951;16:297-334. [doi: 10.1007/BF02289028]
32. Calinski RB, Harabasz J. A dendrite method for cluster analysis. Commun Stat Theory Methods 1974;3(1):1-27.
33. Agresti A. Categorical Data Analysis. 2nd edition. New York, NY: Wiley-Interscience; 2002.
34. Efron B. Computers and the theory of statistics: thinking the unthinkable. SIAM Rev 1979;21:460-480. [doi: 10.1137/1021092]

XSL•FO
RenderX

35. Milczarek M, Schneider E, Gonzalez ER. OSH in Figures: Stress at Work, Facts and Figures. Luxembourg: European Communities; 2009.
36. Hjemdahl P. Stress and the metabolic syndrome: an interesting but enigmatic association. Circulation 2002 Nov 19;106(21):2634-2636 [FREE Full text] [Medline: 12438283]
37. Siegrist J. Adverse health effects of high-effort/low-reward conditions. J Occup Health Psychol 1996 Jan;1(1):27-41. [Medline: 9547031]
38. Karasek R, Baker D, Marxer F, Ahlbom A, Theorell T. Job decision latitude, job demands, and cardiovascular disease: a prospective study of Swedish men. Am J Public Health 1981 Jul;71(7):694-705 [FREE Full text] [Medline: 7246835]
39. Rozanski A. Integrating psychologic approaches into the behavioral management of cardiac patients. Psychosom Med 2005;67 Suppl 1:S67-S73 [FREE Full text] [doi: 10.1097/01.psy.0000164252.07368.81] [Medline: 15953806]
40. Hamer M, Molloy GJ, Stamatakis E. Psychological distress as a risk factor for cardiovascular events: pathophysiological and behavioral mechanisms. J Am Coll Cardiol 2008 Dec 16;52(25):2156-2162. [doi: 10.1016/j.jacc.2008.08.057] [Medline: 19095133]
41. Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. JAMA 2002 Jan 16;287(3):356-359 [FREE Full text] [Medline: 11790215]
42. Goldberg SI, Niemierko A, Turchin A. Analysis of data errors in clinical research databases. AMIA Annu Symp Proc 2008:242-246. [Medline: 18998889]
43. Narayan KM, Aviles-Santa L, Oza-Frank R, Pandey M, Curb D, McNeely M, Cardiovascular Disease in Asian and Pacific Islander Populations NHLBI Working Group. Report of a National Heart, Lung, And Blood Institute Workshop: heterogeneity in cardiometabolic risk in Asian Americans In the U.S. Opportunities for research. J Am Coll Cardiol 2010 Mar 9;55(10):966-973. [doi: 10.1016/j.jacc.2009.07.075] [Medline: 20202512]
44. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. Am Psychol 2004;59(2):93-104. [doi: 10.1037/0003-066X.59.2.93] [Medline: 14992636]
45. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008 Feb 12;117(6):743-753 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.699579] [Medline: 18212285]
46. Graham I, Atar D, Borch-Johnsen K, Boysen G, Burell G, Cifkova R, European Society of Cardiology (ESC), European Association for Cardiovascular Prevention and Rehabilitation (EACPR), Council on Cardiovascular Nursing, European Association for Study of Diabetes (EASD), International Diabetes Federation Europe (IDF-Europe), European Stroke Initiative (EUSI), International Society of Behavioural Medicine (ISBM), European Society of Hypertension (ESH), European Society of General Practice/Family Medicine (ESGP/FM/WONCA), European Heart Network (EHN). European guidelines on cardiovascular disease prevention in clinical practice: executive summary. Fourth Joint Task Force of the European Society of Cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). Eur J Cardiovasc Prev Rehabil 2007 Sep;14 Suppl 2:E1-40. [doi: 10.1097/01.hjr.0000277984.31558.c4] [Medline: 17726406]
47. Svetkey LP. Management of prehypertension. Hypertension 2005 Jun;45(6):1056-1061 [FREE Full text] [doi: 10.1161/01.HYP.0000167152.98618.4b] [Medline: 15897368]
48. Lien LF, Brown AJ, Ard JD, Loria C, Erlinger TP, Feldstein AC, et al. Effects of PREMIER lifestyle modifications on participants with and without the metabolic syndrome. Hypertension 2007 Oct;50(4):609-616 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.107.089458] [Medline: 17698724]
49. Jonas S, Phillips EM. ACSM's Exercise is Medicine: A Clinician's Guide to Exercise Prescription. Philadelphia, PA: Lippincott Williams & Wilkins; 2009.
50. Okie S. The employer as health coach. N Engl J Med 2007 Oct 11;357(15):1465-1469. [doi: 10.1056/NEJMp078152] [Medline: 17928592]

Abbreviations

4SQ: Somatic Stress Symptoms Score Questionnaire
ANOVA: analysis of variance
APR: adjusted Pearson residual
ATP III: Adult Treatment Panel III
CH: Calinski-Harabasz
IT: information technology
MANOVA: multivariate analysis of variance
MeS: metabolic syndrome
MeTs: metabolic equivalents
PCA: principal component analysis
PRINCALS: nonlinear principal component analysis

XSL•FO
RenderX

VAF: variance accounted for

Edited by G Eysenbach; submitted 10.03.11; peer-reviewed by J Dickerson, M Ivannikov; comments to author 09.05.11; revised version received 09.07.11; accepted 26.08.11; published 08.11.11.

Please cite as:
Lucini D, Solaro N, Lesma A, Gillet VB, Pagani M
Health Promotion in the Workplace: Assessing Stress and Lifestyle With an Intranet Tool
J Med Internet Res 2011;13(4):e88
URL: http://www.jmir.org/2011/4/e88/
doi:10.2196/jmir.1798
PMID:22068357

©Daniela Lucini, Nadia Solaro, Alessandro Lesma, Veronique Bernadette Gillet, Massimo Pagani. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improvement of Physical Activity by a Kiosk-based Electronic Screening and Brief Intervention in Routine Primary Health Care: Patient-Initiated Versus Staff-Referred

Matti Leijon[1,2*], MPH, PhD; Daniel Arvidsson[1*], MSc, PhD; Per Nilsen[2], PhD; Diana Stark Ekman[3], PhD; Siw Carlfjord[2], MPH; Agneta Andersson[4], PhD; Anne Lie Johansson[2], BS; Preben Bendtsen[2,5], MD, PhD

[1]Center for Primary Health Care Research, Lund University/Region Skåne, Malmö, Sweden

[2]Department of Medical and Health Sciences, Division of Community Medicine, Linköping University, Linköping, Sweden

[3]Department of Nursing, Health and Culture, University West, Trollhättan, Sweden

[4]County Council of Östergötland, R&D Department of Local Health Care, Linköping University, Linköping, Sweden

[5]Motala Hospital, Department of Medicine, Motala, Sweden

[*]these authors contributed equally

Corresponding Author:
Matti Leijon, MPH, PhD
Center for Primary Health Care Research
Lund University/Region Skåne
CRC, building 28, floor 11, entrance 72, SUS Malmö
Malmö, 205 02
Sweden
Phone: 46 40 391 379
Fax: 46 40 391 370
Email: matti.leijon@med.lu.se

Abstract

Background: Interactive behavior change technology (eg, computer programs, Internet websites, and mobile phones) may facilitate the implementation of lifestyle behavior interventions in routine primary health care. Effective, fully automated solutions not involving primary health care staff may offer low-cost support for behavior change.

Objectives: We explored the effectiveness of an electronic screening and brief intervention (e-SBI) deployed through a stand-alone information kiosk for promoting physical activity among sedentary patients in routine primary health care. We further tested whether its effectiveness differed between patients performing the e-SBI on their own initiative and those referred to it by primary health care staff.

Methods: The e-SBI screens for the physical activity level, motivation to change, attitudes toward performing the test, and physical characteristics and provides tailored feedback supporting behavior change. A total of 7863 patients performed the e-SBI from 2007 through 2009 in routine primary health care in Östergötland County, Sweden. Of these, 2509 were considered not sufficiently physically active, and 311 of these 2509 patients agreed to participate in an optional 3-month follow-up. These 311 patients were included in the analysis and were further divided into two groups based on whether the e-SBI was performed on the patient´s own initiative (informed by posters in the waiting room) or if the patient was referred to it by staff. A physical activity score representing the number of days being physically active was compared between baseline e-SBI and the 3-month follow-up. Based on physical activity recommendations, a score of 5 was considered the cutoff for being sufficiently physically active.

Results: In all, 137 of 311 patients (44%) were sufficiently physically active at the 3-month follow-up. The proportion becoming sufficiently physically active was 16/55 (29%), 40/101 (40%), and 81/155 (52%) for patients with a physical activity score at baseline of 0, 1 to 2, and 3 to 4, respectively. The patient-initiated group and staff-referred group had similar mean physical activity scores at baseline (2.1, 95% confidence interval [CI] 1.8-2.3, versus 2.3, 95% CI 2.1-2.5) and at follow-up, (4.1, 95% CI 3.4-4.7, vs 4.2, 95% CI 3.7-4.8).

Conclusions: Among the sedentary patients in primary health care who participated in the follow-up, the e-SBI appeared effective at promoting short-term improvement of physical activity for about half of them. The results were similar when the

XSL•FO
RenderX

e-SBI was patient-initiated or staff-referred. The e-SBI may be a low-cost complement to lifestyle behavior interventions in routine primary health care and could work as a stand-alone technique not requiring the involvment of primary health care staff.

(J Med Internet Res 2011;13(4):e99) doi:10.2196/jmir.1745

KEYWORDS

Computer-tailored; eHealth; lifestyle behavior; exercise; automated

Introduction

Physical inactivity is acknowledged to be the fourth leading risk factor for global mortality [1]. In Sweden, it has been estimated that physical inactivity contributes to 3.5% of the burden of disease [2]. Hence, effective intervention methods are needed to promote a physically active lifestyle in the population.

Primary health care has been acknowledged as a strategic setting for lifestyle behavior interventions, as indicated by the rapid increase in the number of studies in this field during the last decade [3-6]. There is evidence of both short-term [3-5] and long-term [6] increases in physical activity following counseling provided in primary health care. However, most of these studies were not performed as part of routine care, and they often involved additional personnel and/or patient contacts to those that are usually available. Several barriers to the implementation of lifestyle behavior interventions in routine care have been discussed in the literature, such as insufficient time, high costs, lack of financial reimbursement, perceptions of poor patient adherence to the interventions, limited confidence in providing counseling on lifestyle behaviors, insufficient knowledge about the benefits of physical activity, and lack of appropriate tools to assess and prescribe physical activity [7].

In the light of these implementation challenges, researchers have suggested that the use of interactive behavior change technology (eg, computer programs, Internet websites, and mobile phones) could facilitate the implementation of lifestyle behaviors interventions in primary health care [3,8]. Such technology may address at least some of the barriers to offering face-to-face lifestyle behavior interventions, such as high intervention costs, lack of time, and lack of knowledge. Although there are numerous studies investigating the effectiveness of computer-based and Web-based interventions, they are rarely performed as part of routine care [9,10].

Acceptability of computer-based interventions has been reported to be high among patients in primary health care [11,12] and to be highest among those who were referred by staff to perform the intervention [11] and whose doctor examined the results [12]. Hence, implementation of computer-based interventions as an integrated part of patient counseling may facilitate the use of such interventions in lifestyle behavior change. On the other hand, involvement of primary health care staff in computer-based interventions increases intervention costs, which is important because one of the concepts behind these techniques is to provide automated, stand-alone support to lifestyle behavior change at a low cost. Hence, we are interested in whether computer-based interventions are effective as stand-alone intervention methods in routine primary health care.

An electronic screening and brief intervention (e-SBI) system has been developed by a research team at Linköping University, Sweden. The system consists of a screening questionnaire collecting lifestyle data and an immediate feedback system that reports patient risk level and provides tailored advice for lifestyle behavior change. The e-SBI can be set up to be performed as part of ordinary patient counseling in primary health care or as stand-alone computer stations with touch screens without staff referral. Since it was started in the fall of 2006, the e-SBI has been successively implemented in primary health care in Östergötland County, Sweden. Results describing different aspects of the implementation phase have been reported previously [11,13-15]. We have since begun to evaluate the effectiviness of the e-SBI. We started by focusing on differences between patient-initiated and staff-referred e-SBIs. The initial results of this evaluation for behavior change concerning alcohol intake have recently been published [16]. They showed that the e-SBI had a positive influence on alcohol consumption that did not differ according to whether it was patient-initiated or staff-referred. In the present study, we explored the effectiveness of the physical activity module of the e-SBI and whether it differed between patients who performed the e-SBI on their own initiative and those who were referred to it by primary health care staff.

Methods

Study Location and Patients

The study was conducted in Östergötland County, Sweden, which had approximately 420,000 inhabitants during the study period (2007-2009). There were 42 operating primary health care units within the county when the study was performed. The units differed with regard to number of listed patients aged 18 years and over (average 9500, range 4200 to 16,500) and the number of general practitioners (GPs), nurses, and other staff members employed.

The number of primary health care units offering patients the e-SBI was successively extended during the study period from 10 units in 2007 to 28 units in 2009. The included units were situated in both urban and rural areas. The e-SBI was performed anonymously as part of routine health care. Patients performing the e-SBI during a two-year period, from September 2007 to August 2009 and who were not considered to be sufficiently physically active according to the results of the physical activity screening (see the physical activity section below) were included in the study. The patients were further divided into two groups. The first group consisted of patients who performed the e-SBI on their own initiative, hereafter referred to as the *patient-initiated* group. In the second group, hereafter referred to as the *staff-referred* group, the patients were invited to perform the e-SBI after their appointments with primary health

XSL•FO
RenderX

care staff. Referrals were made by GPs, nurses, physiotherapists, or other staff members responsible for consultations involving lifestyle behaviors. Each primary health care unit was allowed to decide who should make the referrals.

The Electronic Screening and Brief Intervention Concept

Primary health care units participating in the study were provided with one or two sets of computers, monitors, and printers depending on enrolled patient population size; all were included in stand-alone, touch-screen information technology (IT) kiosks (Figure 1). The same equipment was used for both patient-initiated and staff-referred tests. It was placed in or close to a waiting room in which a poster providing information about the test was displayed. The e-SBI concept was based on previous findings of using e-SBI in student health care and emergency department settings [17-21]. The e-SBI included health-related questions regarding alcohol consumption, physical activity, motivation to change, and attitudes toward performing the test. A personalized written feedback was received, including summaries on the current physical activity level, and printed out at the kiosk after patients completed the tests. In the present study, only physical activity-related data are presented. A question was included in the e-SBI concerning whether the patient was referred to the test by staff or performed it on his/her own initiative.

Figure 1. The e-SBI touch screen IT kiosk.



XSL•FO
RenderX

Physical Activity

The physical activity measure in the e-SBI concept included two separate questions based on the American College of Sports Medicine/American Heart Association recommendation from 2007 [22]. This recommendations says that adults should reach 5 days of 30 minutes of moderate activity, 3 days of 20 minutes of vigorous activity, or a combination of both (eg, walking briskly for 30 minutes on 2 days during a week and jogging for 20 minutes on 2 other days). Hence, the first question in the e-SBI asked the participants the number of days in a usual week they performed at least 30 minutes (in bouts of at least 10 minutes each) of moderate physical activity, and the second asked the participants the number of days in a usual week they performed at least 20 minutes of vigorous physical activity. A physical activity score was calculated by summing the number of days during which the required amount of moderate or vigorous physical activity was performed. On occasions of combinations of days of moderate and vigorous physical activity, the number of days of moderate physical activity was weighted by 1 and vigorous physical activity by 1.7 (5 days/3 days = 1.7) when calculating the physical activity score. The physical activity score could have a value between 0 and 18.9, and the cutoff for sufficient physical activity set to 5 (fulfilling the American College of Sports Medicine/American Heart Association recommendation).

Follow-up at 3 Months

After completing an e-SBI but before receiving a personalized printout, each patient was invited to participate in an optional follow-up mail survey 3 months later. Those who accepted this invitation were asked to register their national identification number at the end of the test, and they received a questionnaire by mail 3 months later. Mail addresses were retrieved from the Swedish population register. The mailed questionnaire included the same questions about moderate and vigorous physical activity as used at baseline. A reminder was sent 2 weeks after the follow-up questionnaire to those who had not returned the questionnaire.

Based on their response to the invitation to participate in the follow-up mail survey and completion of the follow-up questionnaire, patients were further categorized into three groups: *nonparticipants* completed the e-SBI but did not agree to participate in the follow-up survey, *nonresponders* completed the e-SBI and agreed to participate in the follow-up survey but did not respond to the follow-up questionnaire, and *responders* provided information at both baseline and follow-up.

Ethics

Since the data collection was performed as part of routine health care and the data consisted only of responses to a written questionnaire provided by patients who had given informed consent, there was, according to Swedish law, no need for formal ethical approval at the time at which the data collection was started. However, since then—in June 2008—the regulations were changed due to uncertainty about how to distinguish between routine and research data collection. For new studies involving similar data collection methods, ethical approval would now be required.

Statistics

Baseline data from nonparticipants, nonresponders, and responders were compared to determine the representativeness of participants in the follow-up (responders). Pearson's chi-square test was used to analyze differences in terms of sociodemographic characteristics. Also, mean (95% confidence interval [CI]) and median (interquartile range [IQR]) physical activity scores were compared (Tables 1).

Among responders, physical activity score and physical activity score category at baseline and 3-month follow-up were compared between patients who performed the e-SBI on their own initiative and those who were referred to it by primary health care staff. Created were four physical activity score categories: 0, 1 to 2, 3 to 4, and greater than or equal to 5. Pearson's chi-square test was used to analyze differences in physical activity score category at baseline and at 3-month follow-up, together with comparison of mean (95% CI) and median (IQR) physical activity scores (Table 3). Improvement in physical activity by the e-SBI was assessed from the physical activity score and physical activity score category at follow-up compared with baseline. All statistical analyses were performed using SPSS 18.0 (SPSS Inc, Chicago, IL, USA).

Results

Participation

A total of 7863 patients completed the e-SBI during the two-year sampling period (Figure 2). Of these, 2509 were categorized as not being sufficiently physically active (physical activity score < 5) and were included in the study. Among the included patients, more performed the e-SBIs on their own initiative (1602/2509 or 64%) than were referred to it by primary health care staff (907/2509 or 36%). However, the proportion of patients agreeing to participate in the follow-up was larger in the staff-referred group than in the patient-initiated group at 34% (305/907) versus 13% (208/1602). The final proportion of patients who completed the follow-up (responders) was 20% and 8% in the staff-referred and patient-initiated groups, respectively.

In the patient-initiated group, the proportion of older patients at baseline was significantly higher among responders compared with nonresponders and nonparticipants. However, there were no significant differences in gender distribution or physical activity score among the three groups (Table 1).

In the staff-referred group, the proportion of men was significantly higher among nonresponders compared with the other groups (Table 2). Also, the proportion of older patients was significantly higher among responders compared with nonparticipants. There was, however, no difference in physical activity score among the three groups.

XSL•FO
RenderX

Table 1. Patient-initiated e-SBI: baseline characteristics of nonparticipants, nonresponders, and responders

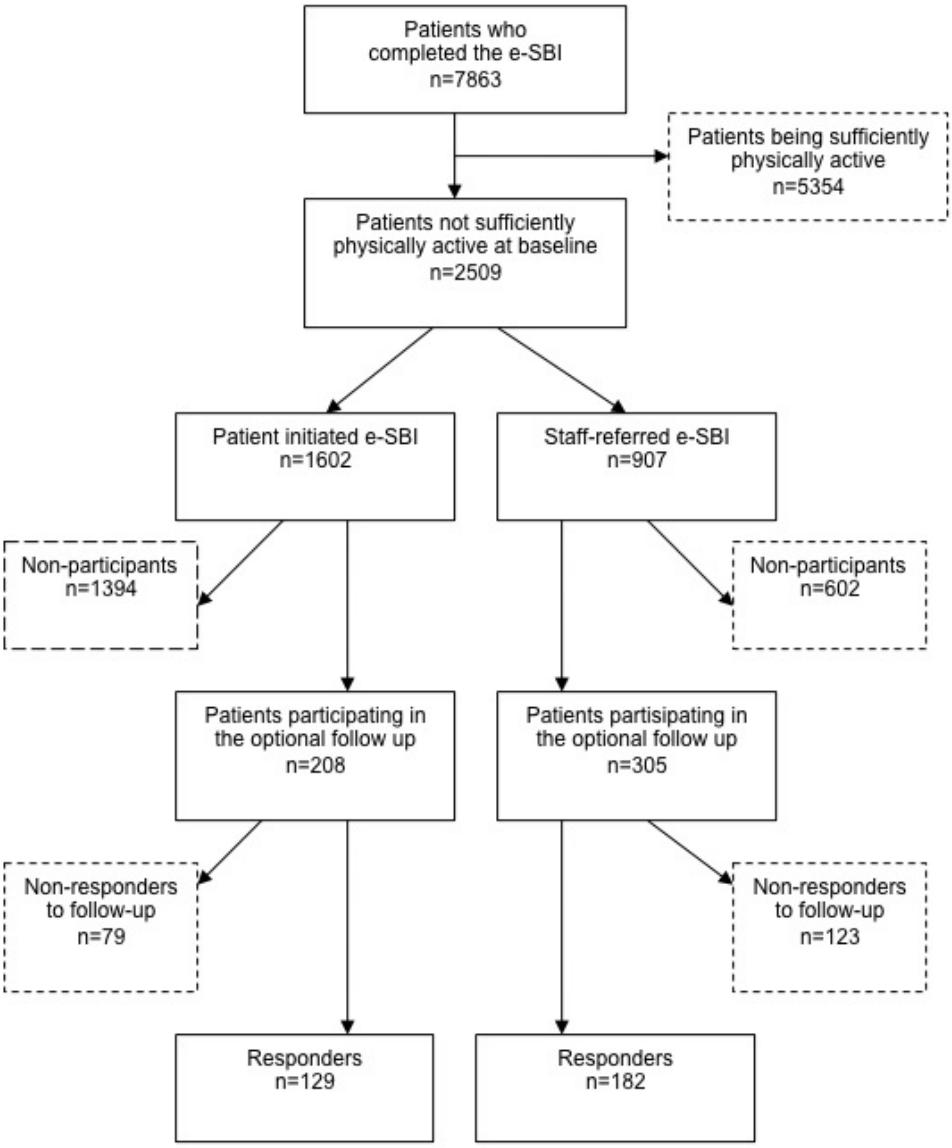
	Nonparticipants	*P* Value, Nonparticipants vs Nonresponders	Nonresponders	*P* Value, Nonresponders vs Responders	Responders	*P* Value, Nonparticipants vs Responders
Gender, n (%) (*P* = .09)						
Men	716 (51)		35 (44)		55 (43)	
Women	678 (49)		44 (56)		74 (57)	
Total	1394 (100)	.25	79 (100)	.89	129 (100)	.07
Age, n (%) (*P* < .001)						
18-20	125 (9)		10 (13)		3 (2)	
21-30	233 (17)		19 (24)		18 (14)	
31-40	360 (26)		23 (29)		22 (17)	
41-50	237 (17)		6 (8)		12 (9)	
51-60	216 (16)		10 (13)		31 (24)	
≥ 61	211 (15)		11 (14)		42 (33)	
Total	1382 (100)	.15	79 (100)	< .001	128 (100)	< .001
Physical activity score						
Mean (95% CI)	1.9 (1.8-2.0)		1.8 (1.5-2.2)		2.1 (1.8-2.3)	
Median (IQR)	2 (0-3)		2 (0-3)		2 (1-3)	

Table 2. Staff-referred e-SBI: baseline characteristics of nonparticipants, nonresponders and responders

	Nonparticipants	*P* Value, Nonparticipants vs Nonresponders	Nonresponders	*P* Value, Nonresponders vs Responders	Responders	*P* Value, Nonparticipants vs Responders
Gender, n (%) (*P* = .06)						
Men	319 (53)		78 (63)		92 (51)	
Women	283 (47)		45 (37)		90 (50)	
Total	602 (100)	.04	123 (100)	.03	182 (100)	.61
Age, n (%)(*P*= .001)						
18-20	54 (9)		4 (3)		5 (3)	
21-30	72 (12)		12 (10)		11 (6)	
31-40	82 (14)		16 (13)		21 (12)	
41-50	108 (18)		27 (22)		26 (14)	
51-60	123 (21)		35 (29)		56 (31)	
≥ 61	156 (26)		28 (23)		63 (35)	
Total	595 (100)	.11	122 (100)	.18	182 (100)	< .001
Physical activity score						
Mean (95% CI)	2.0 (1.9-2.2)		2.1 (1.8-2.3)		2.3 (2.1-2.5)	
Median (IQR)	2 (1-3)		2 (1-3)		3 (1-3)	

XSL•FO
RenderX

Figure 2. Flowchart of the recruitment of patients.



Improvement in Physical Activity (Responders Only)

There was no statistically significant difference in physical activity score or physical activity score category between the patient-initiated and staff-referred groups at baseline or at the 3-month follow-up (Table 3). For all responders together, mean physical activity score was significantly higher at 3-month follow-up (4.2, 95% CI 3.8-4.6) compared with baseline (2.2, 95% CI 2.0-2.3). The median (IQR) score increased only slightly, from 2.5 (1-3) to 3 (1-6). However, 44% (136/311) of the patients succeded in becoming sufficiently physical active at the 3-month follow-up.

XSL•FO
RenderX

Table 3. Physical activity score and category distribution at baseline and follow-up

Physical Activity Score[a]	Patient-Initiated n = 129	Staff-Referred n = 182	All Responders n = 311
Baseline			
Mean (95% CI)	2.1 (1.8-2.3)	2.3 (2.1-2.5)	2.2 (2.0-2.3)
Median (IQR)	2 (1-3)	3 (1-3)	2.5 (1-3)
Physical activity score category, n (%)[b]			
0	24 (19)	31 (17)	55 (18)
1-2	49 (38)	52 (29)	101 (32)
3-4	56 (43)	99 (54)	155 (50)
≥ 5	0 (0)	0 (0)	0 (0)
3 month follow-up			
Mean (95% CI)	4.1 (3.4-4.7)	4.2 (3.7-4.8)	4.2 (3.8-4.6)
Median (IQR)	3 (1-6)	3 (2-6)	3 (1-6)
Physical activity score category, n (%)[c]			
0	20 (16)	21 (12)	41 (13)
1-2	26 (20)	37 (20)	63 (20)
3-4	31 (24)	40 (22)	71 (23)
≥ 5	52 (40)	84 (46)	136 (44)

[a]The physical activity score ranged between 0 and 18.9, and the cutoff for being sufficiently physically active was 5. No patients were categorized as sufficiently physically active at baseline according to inclusion criteria.

[b]$\chi^2_2 = 3.99$ ($P = .14$), patient-initiated versus staff-referred for categories 0, 1-2 and 3-4 at baseline

[c]$\chi^2_2 = 1.63$ ($P = .65$), patient-initiated versus staff-referred for categories 0, 1-2, 3-4 and ≥ 5 at follow-up

Table 4 shows descriptive data of the change in physical activity score category from baseline to follow-up according to physical activity score category at baseline. Of patients with a physical activity score of zero at baseline, 29% (16/55) became sufficiently physically active at the 3-month follow-up. The corresponding proportions for those with physical activity scores of 1 to 2 and 3 to 4 at baseline were 40% (40/101) and 52% (81/155), respectively.

Table 4. Change in physical activity score category from baseline to 3-month follow-up in all responders (n = 311)

		Physical Activity Score Category at 3-Month Follow-up			
Physical activity score category at baseline	n	0 %	1-2 %	3-4 %	≥5 %
0	55	35	20	16	29
1-2	101	13	26	22	40
3-4	155	6	17	26	52

Discussion

In this study, previously sedentary patients in primary health care improved their physical activity 3 months after performing an electronic screening and brief intervention (e-SBI). Overall, 44% of the patients became sufficiently physically active and the improvement in physical activity was similar when the e-SBI was patient-initiated or staff-referred.

These results suggest that the e-SBI is an effective method for promoting a short-term increase in physical activity in patients in primary health care. The e-SBI may be employed as a part of routine care, but there are several factors that need to be taken into account for implementation to be effective. These include staff expectations, perceived need for the innovation to be implemented, compatibility with existing routines, and implementation strategy [11,13-15]. The e-SBI may be used as an integrated part of lifestyle behavior counseling, which may promote greater patient acceptability [11,12] and adherence to the intervention. Patients can be referred by their physician, nurse, or physiotherapist to perform the e-SBI during a visit to the primary health care center. Together, they can then examine the results as part of the consultation, providing background information for physical activity referrals. Alternatively, patients could choose to just bring the printed feedback home for their own use or reference.

XSL•FO
RenderX

The e-SBI could also be used as a stand-alone technique for promoting lifestyle behavior change, as it produces similar results without the involvement of primary health care staff. Posters informing the patients about the e-SBI can be placed in the waiting rooms. This would be an attractive, low-cost option for primary health care. Both patient-initiated and staff-referred e-SBIs may represent cost-effective complements to ordinary face-to-face interventions and may provide sufficient support to those patients who show better acceptance of this kind of technique. This may free up time for patients requiring face-to-face interventions. Although the initial costs of implementating the e-SBI might be high, the running costs, including technical support, would be far less than face-to-face counseling. The e-SBI would, therefore, deliver cost savings in the long run. Besides, the implementation costs for face-to-face interventions, including staff training, may also be high.

In the present study, there was a lower attrition rate at follow-up in the staff-referred group compared with the patient-initiated group. The extra attention/support experienced by the staff-referred group may have promoted continued participation in the study. However, this does not mean that all patients in the staff-referred group had sufficient motivation to improve their physical activity on their own. In the patient-initiated group, patients who remained in the study may have been those with higher internal motivation for behavior change. Hence, the extra attention/support provided to the staff-referred group versus the motivational characteristics of the remaining participants in the patient-initiated group may explain similarities in improvement in physical activity between the groups. The e-SBI (patient-initiated or staff-referred) may be adapted to meet the support needs of individual patients.

The results of the present study can be compared with those of our previous study of the effect of physical activity referral in routine primary health care in Östergötland County [23]. A typical physical activity referral was performed as a face-to-face counseling session by a physician, nurse, or physiotherapist and resulted in a physical activity prescription. Physical activity was followed up after 3 months through a telephone interview, a postal questionnaire, or during a normal return visit. Participants were asked to give the number of days of at least 30 minutes of physical activity of at least moderate intensity during a week [23]. In the present study, 44% of the patients were sufficiently physically active at follow-up. The proportions for patients who were physically active on 0, 1 to 2, and 3 to 4 days per week at baseline were 29%, 40%, and 52%, respectively. The corresponding proportions in our previous study were 29% for all patients, and 26%, 25%, and 40% for patients who were physically active on 0, 1 to 2, and 3 to 4 days per week at baseline, respectively [23]. The differences in proportions between the studies may partly be explained by the different questions used, although both question formats aimed to separate those who reached the recommended activity level from those who did not. In the present study, the number of days of moderate and vigorous physical activity was assessed by two separate questions, which may have promoted reporting of a higher number of days of physical activity compared with the previous study, in which only one question was used [23]. Also, in the study by Leijon et al, all patients were included in the follow-up [23]. In the present study, the follow-up was optional, which may have caused selection bias through inclusion of the more motivated patients. Considering these methodological differences, one may conclude that the e-SBI promotes short-term improvement in physical activity, similar to physical activity referrals. A study in which the e-SBI and physical activity referrals are directly compared would provide valuable information concerning their complementary roles in routine primary health care.

We are not aware of any comparable e-SBI physical activity study. Carroll et al performed a randomized controlled trial of a theory-based, computerized physical activity intervention in primary health care [24]. However, in their study, physical activity and psychosocial mediators were investigated through mailed surveys, and the responses from the participants were entered into a computer program by research staff. A tailored report generated and designed to motivate and support behavioral change was then mailed back to the participants. There was no significant difference in moderate-to-vigorous physical activity between the intervention group (139 minutes/week) and control group (109 minutes/week) after 6 months of follow-up. The authors reasoned that performing multiple surveys may have caused reactivity to the research protocol, thereby enhancing physical activity. It is known that assessment reactivity can influence intervention outcomes. For instance, Maisto et al showed that less frequent and less comprehensive assessment yielded lower assessment reactivity in a study of alcohol use and alcohol-related consequences [25].

In the first phase of evaluating the effectiviness of the e-SBI in promoting improved behavior concerning physical activity (the present study) and alcohol consumption [16], we have compared the results of patient-initiated and staff-referred e-SBI. A control group representing routine care was not included in the study design. Hence, a limitation of these two stuides is that we are not able to draw any conclusions concerning the effect of complementing routine care with e-SBI. In routine primary health care, it is not always feasible to apply a randomized controlled trial design. However, it may be a necessary complement to the results of our studies and may confirm the effectiveness of the e-SBI in promoting behavior change in routine primary health care. It is also necessary to include longer follow-up periods to determine the long-term effect of the e-SBI.

In the present study, a self-report measure of physical activity was used to assess change in physical activity following the intervention, as it is part of the e-SBI and may be the most feasible way of assessing physical activity in routine primary health care. However, self-report methods suffer from reporting bias, consisting of a combination of reactivity, recall bias, and social desirability [26]. The level of physical activity has been reported to be considerably higher when assessed by self-report methods than when assessed objectively using accelerometers [27]. Objective methods are considered to provide a more accurate measure of physical activity. Hence, the improvement in physical activity in the present study may have been overestimated. Including objective measures of physical activity in future studies may improve our ability to determine the effectiveness of the e-SBI.

XSL•FO
RenderX

There was a large attrition rate at follow-up in the present study, although patients participating in the follow-up were fairly representative of all patients who performed the e-SBI, reducing the risk of selection bias that may otherwise have affected the intervention outcome.

In conclusion, an electronic screening and brief intervention (e-SBI) implemented in routine primary health care improved physical activity for about half of the sedentary patients who agreed to participate in the follow-up. Similar results were obtained when the e-SBI was patient-initiated or staff-referred. The e-SBI may be a low-cost complement to lifestyle behavior interventions in routine primary health care and could work as a stand-alone technique not involving primary health care staff.

Acknowledgments

We are grateful to nutritionist Karolina Krus for providing the computers for the e-SBI to participating primary health care centers and for being responsible for on-site management of study questionnaires and data entry. Financial support was provided by the Swedish National Institute for Public Health.

Conflicts of Interest

Preben Bendtsen is a partner in a company that develops eHealth applications similar to the one described in this paper. The other authors declare no conflicts of interest.

References

1. World Health Organization. Global Recommendations on Physical Activity for Health. Geneva, Switzerland: WHO Press, World Health Organization; 2010.
2. Agardh E, Moradi T, Allebeck P. [The contribution of risk factors to the burden of disease in Sweden. A comparison between Swedish and WHO data]. Lakartidningen 2008;105(11):816-821. [Medline: 18461873]
3. Eakin EG, Glasgow RE, Riley KM. Review of primary care-based physical activity intervention studies: effectiveness and implications for practice and future research. J Fam Pract 2000 Feb;49(2):158-168. [Medline: 10718694]
4. Lawlor DA, Hanratty B. The effect of physical activity advice given in routine primary care consultations: a systematic review. J Public Health Med 2001 Sep;23(3):219-226 [FREE Full text] [Medline: 11585195]
5. Smith BJ. Promotion of physical activity in primary health care: update of the evidence on interventions. J Sci Med Sport 2004 Apr;7(1 Suppl):67-73. [Medline: 15214604]
6. Müller-Riemenschneider F, Reinhold T, Nocon M, Willich SN. Long-term effectiveness of interventions promoting physical activity: a systematic review. Prev Med 2008 Oct;47(4):354-368. [doi: 10.1016/j.ypmed.2008.07.006] [Medline: 18675845]
7. Douglas F, Torrance N, van Teijlingen E, Meloni S, Kerr A. Primary care staff's views and experiences related to routinely advising patients about physical activity. A questionnaire survey. BMC Public Health 2006;6:138 [FREE Full text] [doi: 10.1186/1471-2458-6-138] [Medline: 16719900]
8. Whitlock EP, Orleans CT, Pender N, Allan J. Evaluating primary care behavioral counseling interventions: an evidence-based approach. Am J Prev Med 2002 May;22(4):267-284. [Medline: 11988383]
9. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221. [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
10. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
11. Carlfjord S, Nilsen P, Leijon M, Andersson A, Johansson K, Bendtsen P. Computerized lifestyle intervention in routine primary health care: evaluation of usage on provider and responder levels. Patient Educ Couns 2009 May;75(2):238-243. [doi: 10.1016/j.pec.2008.10.004] [Medline: 19046844]
12. Sciamanna CN, Novak SP, Houston TK, Gramling R, Marcus BH. Visit satisfaction and tailored health behavior communications in primary care. Am J Prev Med 2004 Jun;26(5):426-430. [doi: 10.1016/j.amepre.2004.02.003] [Medline: 15165659]
13. Carlfjord S, Lindberg M, Bendtsen P, Nilsen P, Andersson A. Key factors influencing adoption of an innovation in primary health care: a qualitative study based on implementation theory. BMC Fam Pract 2010;11:60 [FREE Full text] [doi: 10.1186/1471-2296-11-60] [Medline: 20731817]
14. Carlfjord S, Andersson A, Nilsen P, Bendtsen P, Lindberg M. The importance of organizational climate and implementation strategy at the introduction of a new working tool in primary health care. J Eval Clin Pract 2010 Dec;16(6):1326-1332. [doi: 10.1111/j.1365-2753.2009.01336.x] [Medline: 20738475]
15. Carlfjord S, Andersson A, Bendtsen P, Nilsen P, Lindberg M. Applying the RE-AIM framework to evaluate two implementation strategies used to introduce a tool for lifestyle intervention in Swedish primary health care. Health Promot Int 2011 Mar 11. [doi: 10.1093/heapro/dar016] [Medline: 21398336]

XSL•FO
RenderX

16. Bendtsen P, Stark Ekman D, Johansson A, Carlfjord S, Andersson A, Leijon M, et al. Referral to an electronic screening and brief alcohol intervention in primary health care in sweden: impact of staff referral to the computer. Int J Telemed Appl 2011;2011:918763. [doi: 10.1155/2011/918763] [Medline: 21603024]
17. Andersson A, Wiréhn AB, Olvander C, Ekman DS, Bendtsen P. Alcohol use among university students in Sweden measured by an electronic screening instrument. BMC Public Health 2009;9:229 [FREE Full text] [doi: 10.1186/1471-2458-9-229] [Medline: 19594906]
18. Nilsen P, Festin K, Guldbrandsson K, Carlfjord S, Holmqvist M, Bendtsen P. Implementation of a computerized alcohol advice concept in routine emergency care. Int Emerg Nurs 2009 Apr;17(2):113-121. [doi: 10.1016/j.ienj.2008.11.006] [Medline: 19341997]
19. Trinks A, Festin K, Bendtsen P, Nilsen P. Alcohol consumption and motivation to reduce drinking among emergency care patients in Sweden. Int J Inj Contr Saf Promot 2009 Sep;16(3):133-141. [doi: 10.1080/17457300903024087] [Medline: 19941211]
20. Bendtsen P, Holmqvist M, Johansson K. Implementation of computerized alcohol screening and advice in an emergency department--a nursing staff perspective. Accid Emerg Nurs 2007 Jan;15(1):3-9. [doi: 10.1016/j.aaen.2006.09.004] [Medline: 17113773]
21. Karlsson A, Bendtsen P. Acceptability of a computerized alcohol screening and advice routine in an emergency department setting--a patient perspective. Addict Behav 2005 May;30(4):767-776. [doi: 10.1016/j.addbeh.2004.08.022] [Medline: 15833580]
22. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, American College of Sports Medicine, American Heart Association. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Circulation 2007 Aug 28;116(9):1081-1093 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.185649] [Medline: 17671237]
23. Leijon ME, Bendtsen P, Nilsen P, Festin K, Ståhle A. Does a physical activity referral scheme improve the physical activity among routine primary health care patients? Scand J Med Sci Sports 2009 Oct;19(5):627-636. [doi: 10.1111/j.1600-0838.2008.00820.x] [Medline: 18627557]
24. Carroll JK, Lewis BA, Marcus BH, Lehman EB, Shaffer ML, Sciamanna CN. Computerized tailored physical activity reports. A randomized controlled trial. Am J Prev Med 2010 Aug;39(2):148-156. [doi: 10.1016/j.amepre.2010.04.005] [Medline: 20621262]
25. Maisto SA, Clifford PR, Davis CM. Alcohol treatment research assessment exposure subject reactivity effects: part II. Treatment engagement and involvement. J Stud Alcohol Drugs 2007 Jul;68(4):529-533. [Medline: 17568956]
26. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport 2000 Jun;71(2 Suppl):S1-14. [Medline: 10925819]
27. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the Physical Activity Guidelines for Americans. Am J Prev Med 2011 Apr;40(4):454-461. [doi: 10.1016/j.amepre.2010.12.016] [Medline: 21406280]

Abbreviations

CI: confidence interval
e-SBI: electronic screening and brief intervention
GP: general practitioner

Edited by G Eysenbach; submitted 01.02.11; peer-reviewed by A Booth, R Kretschmann, M Löf; comments to author 12.05.11; revised version received 21.06.11; accepted 11.07.11; published 22.11.11.

Please cite as:
Leijon M, Arvidsson D, Nilsen P, Stark Ekman D, Carlfjord S, Andersson A, Johansson AL, Bendtsen P
Improvement of Physical Activity by a Kiosk-based Electronic Screening and Brief Intervention in Routine Primary Health Care: Patient-Initiated Versus Staff-Referred
J Med Internet Res 2011;13(4):e99
URL: http://www.jmir.org/2011/4/e99/
doi:10.2196/jmir.1745
PMID:22107702

©Matti Leijon, Daniel Arvidsson, Per Nilsen, Diana Stark Ekman, Siw Carlfjord, Agneta Andersson, Anne Lie Johansson, Preben Bendtsen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Web-Based Risk Communication and Planning in an Obese Population: Exploratory Study

Anastasia Soureti[1,2], BSC(Hons), MSc; Peter Murray[1], CStat, Mphil; Mark Cobain[1], PhD; Willem van Mechelen[2], MD; Robert Hurling[1], PhD

[1]Unilever Discover, Colworth Science Park, Bedfordshire, United Kingdom

[2]Department of Public and Occupational Health and EMGO Institute for Health & Care Research, VU University Medical Centre, Amsterdam, Netherlands

Corresponding Author:
Anastasia Soureti, BSC(Hons), MSc
Unilever Discover
Colworth Science Park
Sharnbrook
Bedfordshire, MK44 1LQ
United Kingdom
Phone: 44 1234248189 ext 8189
Fax: 44 1234 24 8010
Email: natasha.soureti@unilever.com

Abstract

Background: A healthy diet, low in saturated fat and high in fiber, is a popular medical recommendation in preventing cardiovascular disease (CVD). One approach to motivating healthier eating is to raise individuals' awareness of their CVD risk and then help them form specific plans to change.

Objectives: The aim was to explore the combined impact of a Web-based CVD risk message and a fully automated planning tool on risk perceptions, intentions, and saturated fat intake changes over 4 weeks.

Methods: Of the 1187 men and women recruited online, 781 were randomly allocated to one of four conditions: a CVD risk message, the same CVD risk message paired with planning, planning on its own, and a control group. All outcome measures were assessed by online self-reports. Generalized linear modeling was used to analyze the data.

Results: Self-perceived consumption of low saturated fat foods (odds ratio 11.40, 95% CI 1.86–69.68) and intentions to change diet (odds ratio 21.20, 95% CI 2.6–172.4) increased more in participants allocated to the planning than the control group. No difference was observed between the four conditions with regard to percentage saturated fat intake changes. Contrary to our expectations, there was no difference in perceived and percentage saturated fat intake change between the CVD risk message plus planning group and the control group. Risk perceptions among those receiving the CVD risk message changed to be more in line with their age (change in $slope_{individual}$ = 0.075, P = .01; change in $slope_{comparative}$ = 0.100, P = .001), whereas there was no change among those who did not receive the CVD risk message.

Conclusion: There was no evidence that combining a CVD risk message with a planning tool reduces saturated fat intake more than either alone. Further research is required to identify ways in which matching motivational and volitional strategies can lead to greater behavior changes.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 91154001; http://www.controlled-trials.com/ISRCTN91154001 (Archived by WebCite at http://www.webcitation.org/62sBoGeOO)

(J Med Internet Res 2011;13(4):e100) doi:10.2196/jmir.1579

KEYWORDS

Risk perceptions; cardiovascular disease; planning; saturated fat intake

XSL•FO
RenderX

Introduction

Cardiovascular disease (CVD) is a leading cause of death among adults [1]. A healthy diet, low in saturated fat and high in fiber, is a popular medical recommendation in preventing CVD.

One approach to increase motivation to change is to improve awareness of the risk associated with an unhealthy lifestyle [2,3]. Risk analogies such as Heart-Age (HA) combine aspects of absolute and relative CVD risk and have been found effective in communicating future CVD risk [4,5]. In a recent study, those at higher actual CVD risk who received a HA risk analogy were more aware of their future CVD risk than were those exposed to a percentage CVD risk score [4].

Although many people report having good intentions to eat more healthily, these are not always translated into action [6,7]. Action plans, also known as implementation intentions, are strategies that can bridge the gap between intention and behavior. A meta-analysis of 94 studies showed that implementation intentions had a medium to large effect on goal achievement [8]. Fear appeals may also facilitate change when they are combined with specific instructions on what action to take [9]. While earlier studies explored the value of using action plans [7,10], more recently there has been a greater interest in the characteristics and mechanisms underlying effective plans [11-18], such as the creation of a strong cue–response relationship [12,13].

Research has also investigated the impact of self-efficacy on behavior change. According to the health action process approach model, *action* self-efficacy acts on the motivational part of decision making, whereas *maintenance* self-efficacy acts on the volitional part of the behavior [3]. While some studies report higher self-efficacy in participants making an implementation intention [19,20], others find no difference [7].

Implementation intention research to date has been largely offline (paper and pencil) with little focus and mixed results when their effectiveness has been tested online [21,22]. In a study conducted in an occupational setting, use of online implementation intentions backfired, such that participants who did not form an implementation intention exercised significantly more than participants who formed an implementation intention [21]. In an online dietary intervention, implementation intentions were combined with a text message reminder service leading to a reduction in perceived saturated fat intake and portion sizes [22]. The present study is one of a few studies designed to act on both the motivational and volitional phase of behavior change [23]. We offered a risk communication message to create more appropriate risk perceptions and to increase intention to change, and then helped individuals change their dietary behavior by forming specific plans on how to achieve this. This is also one of the few studies that compared the independent and combined short-term effects of an online health risk communication message and an online implementation intention tool on the promotion of healthy eating in an obese population, who are more likely to be at risk of developing CVD.

Objectives

The primary aim of this investigatory study was to test whether participants could form plans via a fully automated Web-based planning tool (PT) and to assess the short-term effects of combining a CVD risk message (Heart-Age, HA) with the planning tool (HA+PT) on participants' saturated fat intake, measured by a 2-item scale (TIS) and a food frequency questionnaire (FFQ) over a period of 3 weeks. A secondary aim was to assess the effects of the heart-age risk message and planning tool on participants' risk perceptions, self-efficacy, intentions to change saturated fat intake, and intentions to test cholesterol and blood pressure levels. We expected that the heart-age message would primarily change risk perceptions and participants' intentions to change, while the planning tool would act primarily on self-efficacy and behavior. We wanted to explore whether participants could form Web-based plans and whether the combined HA+PT intervention would have a greater impact than either the heart-age message or the planning tool alone.

Methods

Participants

We invited 1187 participants through an online recruitment agency to log in to an open access website to take part in the study. The self-report eligibility criteria included age (30–60 years), obesity (body mass index [BMI] ≥29 kg/m^2), not having a diagnosis of a heart condition or cancer or being pregnant, and being computer savvy. We chose obese participants because they were likely to benefit from heart-health information [24]. To help minimize any imbalance effects created by smokers receiving a higher heart-age score, a UK-representative sample of smokers was distributed across the four conditions of the study.

Design and Procedures

This study was conducted between the middle of January and the end of February 2009 and has been registered retrospectively. It was a Web-based, randomized, between-groups study designed to assess the difference in saturated fat intake between four experimental conditions. No participant–experimenter contact was present. Participants were given online instructions and completed each week's session from the convenience of their home computer. At week 1 (recruitment), participants were recruited by an online agency, signed an online consent form [25], and completed an online questionnaire on their current saturated fat intake, risk perceptions, self-efficacy, and intentions to change their dietary intake. They also received educational information on the importance of a healthy diet low in saturated fat (Multimedia Appendix 1).

At week 2 (intervention), those participants who returned to the website were randomly allocated, using a computer-generated list of random numbers, into one of four conditions: (1) control group (CG), (2) PT condition, (3) HA risk message condition, and (4) HA+PT condition. Allocation of the participants in the four conditions was also stratified to balance by age group (30–45 years or 46–60 years) and gender. In the groups that

XSL•FO
RenderX

received the HA risk message, participants filled out online information on their age, gender, weight, height, prescribed blood pressure medication, family history of heart and vascular disease, smoking status, self-prevalent diabetes, self-reported total and high-density lipoprotein cholesterol levels, and systolic blood pressure. They then received feedback on their future CVD risk in the form of the HA risk message. Participants in the PT condition were asked to identify a list of situations in which they would like to change their saturated fat intake and match these situations with a list of behaviors. Participants in all conditions were asked to fill out a shorter version of the questionnaire asked at baseline. At week 2, participants completed the session once and were not able to revisit the website to make any changes (eg, to create more plans).

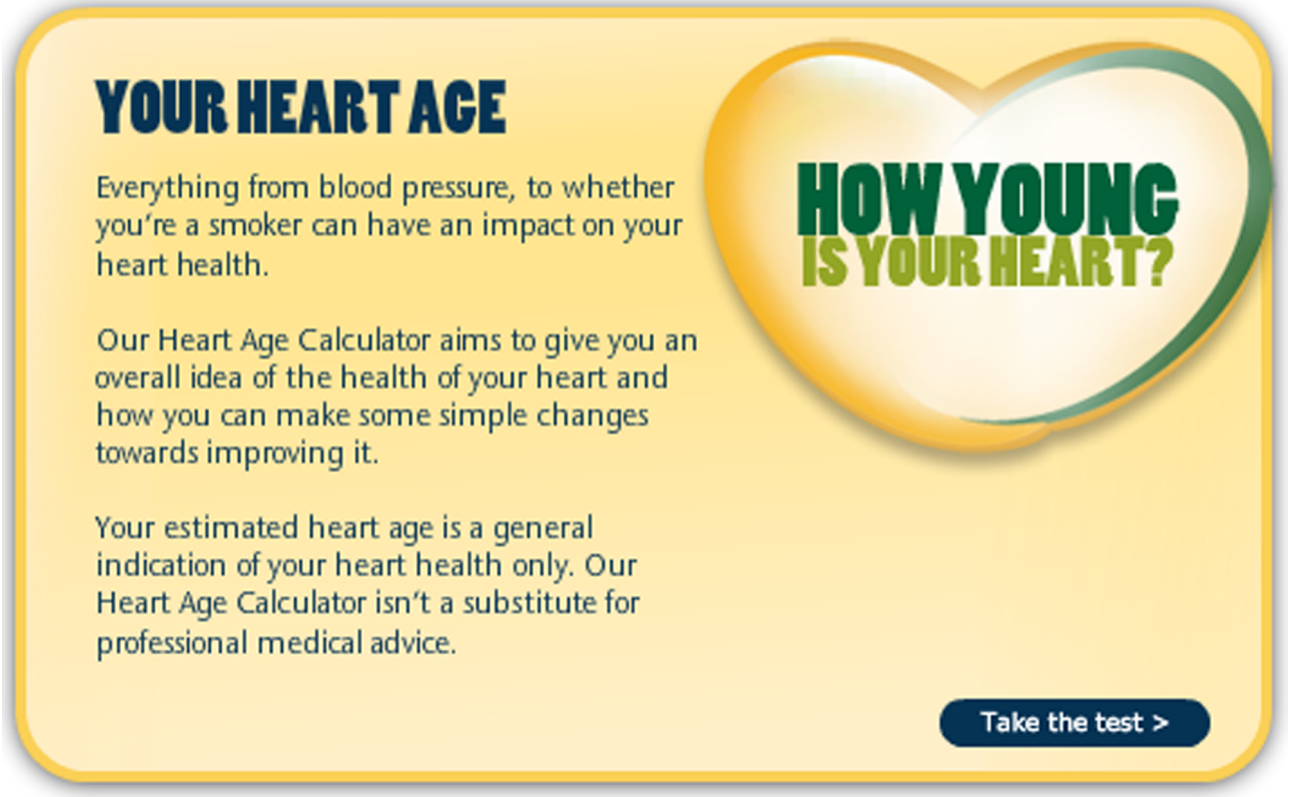
At week 5 (follow-up), participants were asked to complete a follow-up assessment. They received £15 on study completion and were entered in a prize draw for vouchers (£200).

Interventions

The Heart-Age Risk Message Condition

Heart-age, which is described in more detail elsewhere [26], is the age corresponding to someone of the same gender with the same CVD risk level but with normal risk factors. The definition of normal is based on the following profile: not smoking, not diabetic, systolic blood pressure 125 mmHg (midpoint of normal range: 120–130 mmHg), total serum cholesterol 180 mg/dL (4.66 mmol/L; between normal range of 160–200 mg/dL or 4.14–5.18 mmol/L), and high-density lipoprotein cholesterol 45 mg/dL (1.17 mmol/L). For example, a 61-year-old man who smokes and has no other risk factors has a 10-year CVD risk of 10% and the HA of a 73-year-old man. In the HA condition, users filled in an online questionnaire and received feedback in form of the HA risk message (Figure 1).

Figure 1. The Heart-Age risk message.



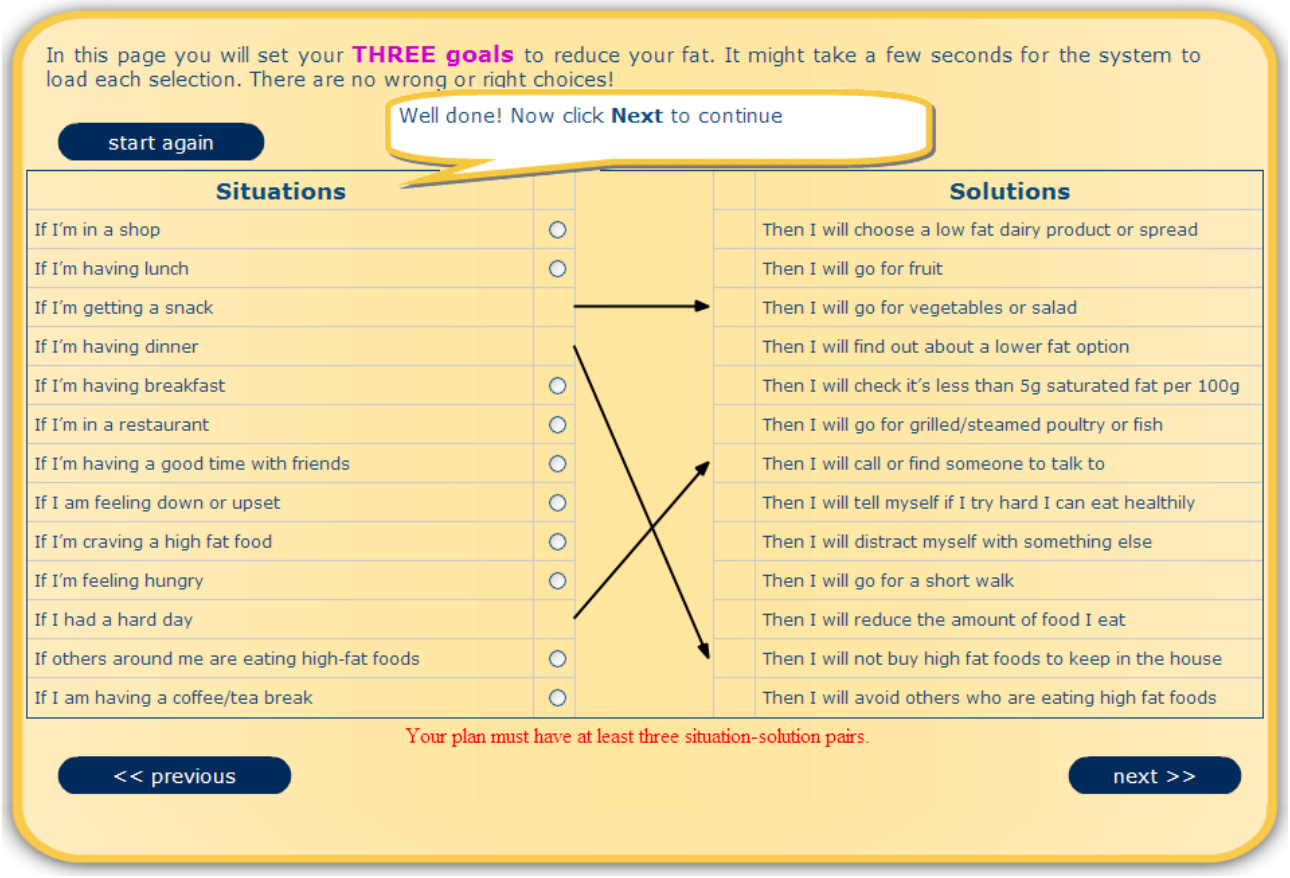
The Planning Tool Condition

Participants who received the PT selected from a list of 13 situations, in which they were tempted to eat unhealthily and then chose an approach to change their behavior from a list of 13 solutions. For every situation–solution pair chosen, a line was drawn visually linking the two together [27]. Participants were asked to complete at least 3 situation–solution pairs.

The solutions were based on constructs from the processes of change model (eg, counterconditioning, stimulus control, and helpful relationships) [28]. Some nutritionally based behaviors were also included from an accredited site [29] after review by an expert nutritionist. The list of situations consisted of both situational cues (eg, having lunch) focusing on the "when and where" and motivational cues (eg, feeling bored) linked to the reasons ("why") for performing a specific behavior [30]. Motivational cues were divided into three main situations: (1) experiencing positive affect, (2) experiencing negative affect, and (3) being faced with cravings [31,32]. The situations were translated into "if" statements (eg, "If I'm having breakfast") and the list of solutions was translated into "then" statements (eg, "then I will tell myself I can eat healthily"). Figure 2 shows the PT.

XSL•FO
RenderX

Figure 2. The planning tool.



Control Group Condition

Participants in the CG received educational information on the importance of a healthy diet low in saturated at week 1 and filled out the same online questionnaires as the rest of the experimental conditions at all study weeks.

Outcome Measures

Saturated fat intake, as the primary outcome measure, was assessed at baseline and follow-up by two measures. First, a self-report index of food [33] was used to record the frequency of consumption of 63 common foods. This FFQ has good test–retest reliability (r = .62, P < .01) [33] and validity when compared with 10-day weighed records [34,35]. Second, a two item scale (TIS) (r = .78, P < .001) was adapted from a previous study [34]. Participants were asked to report their agreement in consumption of low saturated fat foods ("I have eaten foods low in saturated fat...") followed by frequency in consumption of these foods ("How often did you eat foods low in saturated fat?"). The correlation between the two measures was –0.320 (P < .001) at week 1 and –0.291 (P < .001) at week 5. Negative correlations are due to reverse scales used for the self-perceived items.

CVD risk perceptions measured participants' perceived risk in an absolute sense and comparative with their age group [4]. The first item (Q1) examined perceptions of individual CVD risk ("I think that my chances of getting heart disease in the short term are..."). The second item (Q2) compared participants' risk perceptions against those of other people of their age ("Compared to an average person of my age and sex, my chances of getting heart disease are..."). Responses were measured on a 7-point Likert scale at weeks 1 and 2.

Intention to reduce saturated fat intake was measured at weeks 1, 2, and 5 on a 7-point Likert scale via 10 items, which were highly intercorrelated (Cronbach alpha = .92), so were analyzed as a composite score. At follow-up, there were two further questions on participants' intentions to assess their cholesterol and blood pressure over the next month.

Action and maintenance self-efficacy were modified from previous research [3,36-38]. Action self-efficacy (alpha = .84), which was measured at all study times, consisted of 4 items focusing on confidence to overcome obstacles. Maintenance self-efficacy (alpha = .89), assessed only at follow-up, consisted of 11 items exploring confidence in sustaining change in the face of difficulties. Items were measured on a 4-point scale (not at all, barely true, mostly true, exactly true).

Planning and outcome expectancies items were adapted from previous research [3,36-38] and measured on a 4-point scale. Planning comprised 2 items: "I have my own plan regarding (1) when, (2) how to reduce my saturated fat intake." Outcome expectancies consisted of 11 items linked to the positive and negative expectancies of reducing saturated fat intake (eg, "If I reduce my saturated fat intake"... "food won't taste as good," "I will feel good").

Feedback on the intervention was assessed at week 2 and at follow-up on a 7-point Likert scale (ranging from strongly disagree to strongly agree). Participants were asked to rate the intervention in terms of its emotional impact, personal relevance,

XSL•FO
RenderX

interest, trustworthiness, credibility, and enjoyment. All items were adapted from previous studies [39-41].

Statistical Considerations

Analysis of the outcome measures was restricted to those respondents who completed the follow-up assessments. Response to the CVD risk perceptions was analyzed using a generalized linear model with a cumulative logistic link function and multinomial distribution. Baseline scores and heart risk-adjusted age were included as covariates. As with all the analyses other potential covariates (eg, smoking, BMI, social economic status) were retained if significant in the model. Similar models were used for the intention-to-change and intention-to-test questions, and self-efficacy, planning, and feedback items, but omitting the heart risk-adjusted age covariate.

Mean change in self-perceived saturated fat intake within a group was assessed using analysis of variance with baseline included as a covariate. The groups were compared using another generalized linear model with a cumulative logistic link function and multinomial distribution. Data from the index of food was summarized to yield the total calorie intake per participant and the percentage of total energy intake contributed by saturated fat. We analyzed all of these data using analysis of variance models with baseline covariates always included and any other significant covariates retained. All analysis was carried out using version 9.1.3 of SAS software (SAS Institute, Cary, NC, USA).

Local Research Ethical Review Requirement

The study protocol (Multimedia Appendix 1) was approved by an independent research ethics committee (Colworth Research Ethics Committee) in the South of England on December 4, 2008 (Multimedia Appendix 2). All research was conducted in accordance with the Declaration of Helsinki [42].

Results

Participant Baseline Characteristics

At week 1, we invited 1187 people to participate through an online recruitment agency, of whom 1027 completed the initial questionnaire and were invited to take part in the study. At week 2, a total of 781 participants revisited the website and were allocated to one of four conditions. At week 2, 32 of these participants did not complete the online session. At week 5, a total of 581 participants returned to complete the follow-up questionnaire. We excluded 21 participants from the statistical analysis because they did not complete the whole session or due to inaccurate calorie intake reporting (<500 kcal or >5000 kcal per day). The numbers of participants completing each week are shown in Figure 3.

There was no significant difference in percentage saturated fat intake between participants who completed only the week 1 assessment and those who completed the week 5 measures (P = .79). The mean percentage saturated fat intake at week 1 (baseline) was 15.4%, much higher than the UK recommended levels [43]. Table 1 shows participants' baseline characteristics.

Table 1. Participants' baseline characteristics

	Overall	Control group	Planning group	Heart-Age group	Planning plus Heart-Age group	$F_{3,777}$ statistic (P value)
Number	781	195	195	197	194	
Age (years)[a], mean (SD)	46.89 (8.26)	47.05 (8.48)	47.06 (8.11)	46.91 (8.00)	46.56 (8.52)	0.15 (.93)
BMI[b] (kg/m^2)[a], mean (SD)	35.71 (5.71)	35.72 (5.40)	35.51 (6.15)	36.49 (6.42)	35.08 (4.64)	2.10 (.1)
Smokers (%)	25.61	25.13	25.64	25.26	26.40	0.10[c] (.99)

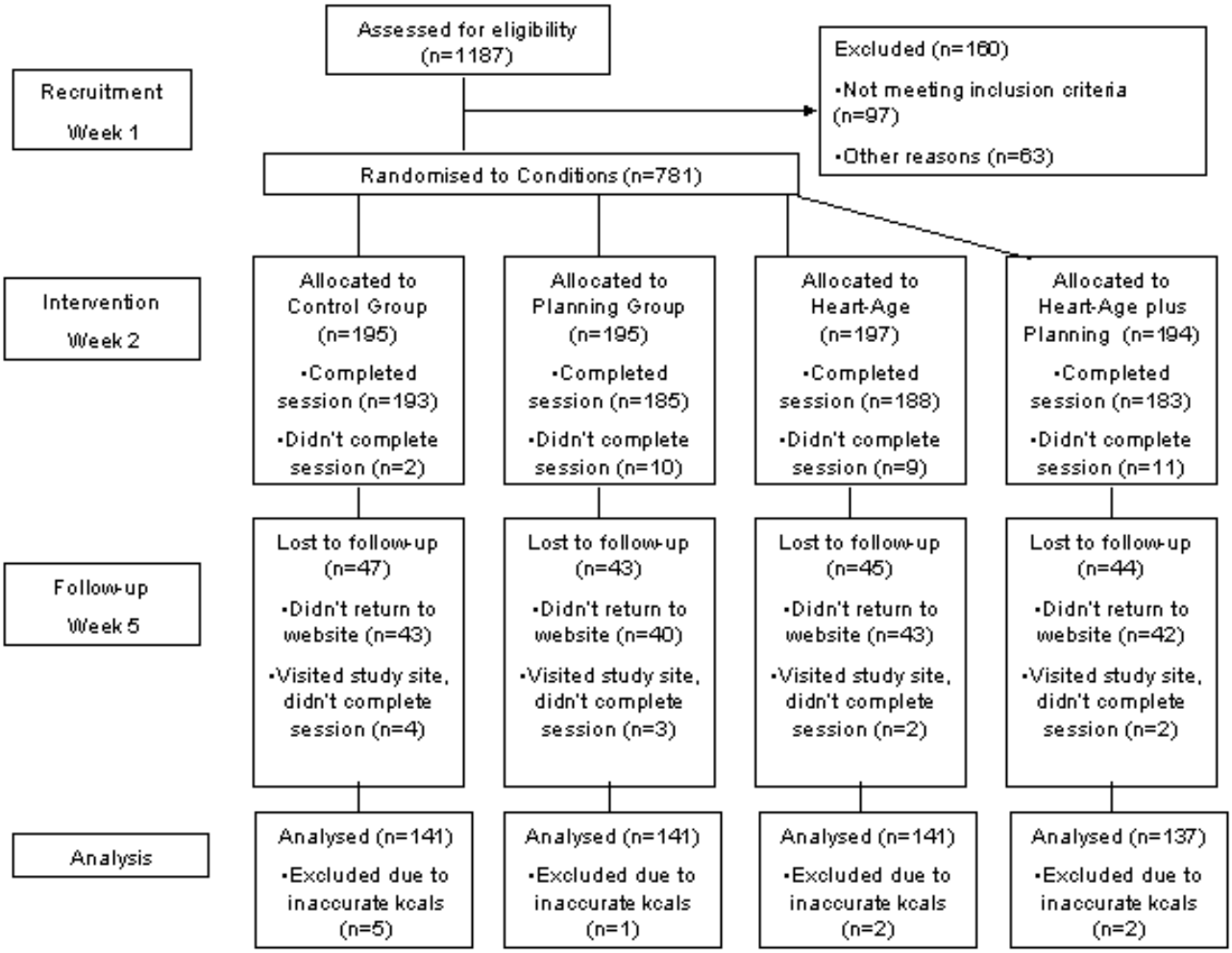
[a] No significant differences found in participants' baseline characteristics (P > .05).

[b] Body mass index.

[c] Chi-square test (c^2_3) statistic and P value.

XSL•FO
RenderX

Figure 3. Flow chart of recruitment, intervention, and follow-up.



Planning Tool

All participants allocated to the PT condition were able to formulate their Web-based plans with an average of 3.9 plans. Participants selected a range of motivational and situational cues. The most frequently chosen situations were “If I’m feeling hungry” (99/747, 13%), “If I’m getting a snack” (97/747, 13%), “If I’m having lunch” (71/747, 10%) “or dinner” (74/747, 10%), “If I’m craving a high-fat food” (66/747, 9%), and “If I’m feeling down or upset” (72/747, 10%).

The most frequently selected solutions were “Then I will go for fruit” (149/747, 19.7%), “Then I will find out about a lower-fat option” (105/747, 14.1%), “Then I will go for grilled/steamed poultry or fish” (85/747, 11%), “Then I will distract myself with something else” (70/747, 9%), and “Then I will tell myself if I try hard I can eat healthily” (66/747, 9%).

Time Spent Online

At week 1, participants spent an average of 12.44 (SD 9.77) minutes online. At week 2, the CG spent the least time online (mean 4.19, SD 2.43 minutes), followed by the PT (mean 7.84, SD 5.18 minutes), the HA (mean 10.91, SD 8.46 minutes), and lastly the HA+PT group (mean 12.47, SD 6.48 minutes). HA+PT spent significantly more time online than the PT group (95% CI, 2.73–6.53) or the CG (95% CI, 6.39–10.18). No significant differences were found at week 2 between the HA+PT and the HA-only condition (95% CI, –0.34 to 3.46). At week 5, there were no further significant differences (P = .67) between the four conditions in time spent filling out the follow-up questionnaire (CG: mean 9.54, SD 3.62 minutes; PT: mean 11.38, SD 9.84 minutes, HA: mean 10.44, SD 5.47 minutes; HA+PT: mean 9.64, SD 6.32 minutes).

Primary Outcomes

Saturated Fat Intake

Participants in all four conditions reported a significant increase in consumption of foods low in saturated fat (the mean of the two self-perceived intake items) between baseline and follow-up, apart from the CG (Table 2). The generalized linear model analysis showed a significant difference between the conditions (c^2_3 = 13.1, P = .005) with respect to perceived saturated fat intake changes. Multiple comparisons of the conditions (with Bonferroni adjustment to allow for the six comparisons) indicated this was due to participants in the PT group reporting a higher perceived increase in low saturated fat foods than those in the CG (odds ratio, 11.40; 95% CI, 1.86–69.68).

XSL•FO
RenderX

Table 2. Saturated fat intake by primary outcome (self-perceived and index of food questionnaire)

Condition	Self-perceived items (baseline mean 4.73)			Index of food questionnaire (baseline mean 15.37%)		
	Week 5	Week 5 – week 1[a]	Pr > |t|	Week 5	Week 5 – week 1[a]	Pr > |t|
Control group	4.857	0.125 (0.100)	.21	14.67	–0.717 (0.198)	.0003
Planning tool	5.087	0.355 (0.102)	.001	14.51	–0.875 (0.198)	<.0001
Heart-Age	4.943	0.212 (0.101)	.04	14.63	–0.748 (0.197)	.0002
Heart-Age + planning tool	4.977	0.245 (0.102)	.02	14.49	–0.893 (0.200)	<.0001

[a] Mean and standard error after adjusting for baseline and other covariates. Note that the standard error for week 5 is the same as the standard error for weeks 5 – 1, due to the use of a baseline covariate in the analysis.

With regard to the index of food, participants in all conditions reported a significant reduction in percentage saturated fat intake between baseline and follow-up (Table 2 with no significant differences found between the four conditions (P = .89).

Secondary Outcomes

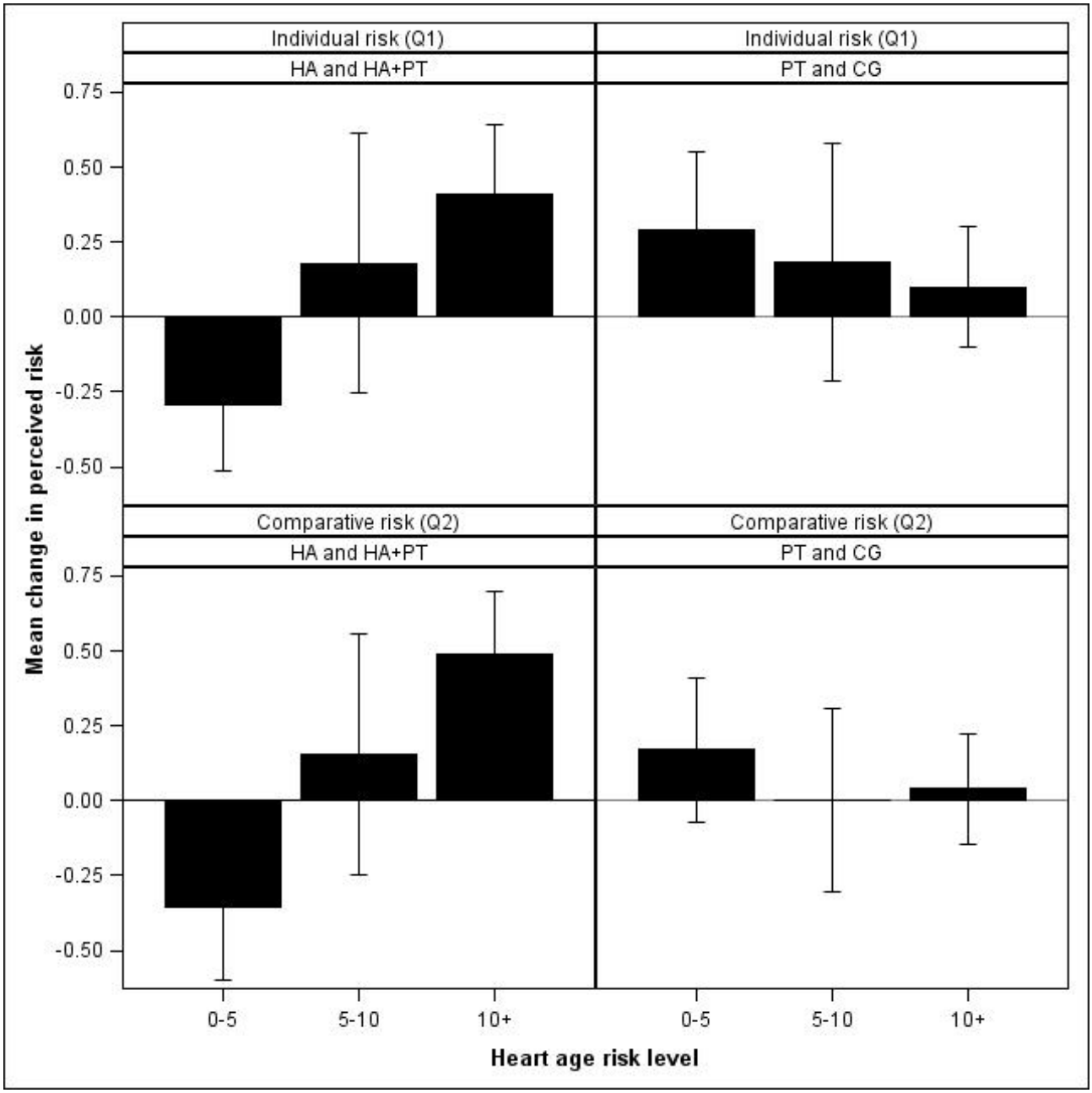
Risk Perceptions

The generalized linear model found no significant differences between the four experimental conditions in terms of their CVD risk perceptions, both for individual (Q1) (P = .88) and comparative risk (Q2) (P = .93). In order to test whether perceived risk was more related to actual risk, we further compared the change in perceived risk between week 2 and week 1 for all participants who received the HA risk message (HA, HA+PT) with those who did not (PT, CG) using a further generalized linear model. Figure 4 shows risk perception changes for Q1 and Q2 split by the different HA risk levels (low: 0–5, moderate: 5–10, and high: 10–15). HA level is the difference between an individual's actual age and his or her risk-adjusted age. For example, the 0–5 HA level includes people whose HA is up to 5 years older than their actual age. For both Q1 and Q2, the regression slopes for those in the HA conditions moved to be more in line with participants' HA risk levels, whereas this was not found for those in the non-HA conditions.

Specifically, for participants in the HA risk message conditions, there was a significant increase in the regression slope of the individual risk perception question (Q1) against the risk-adjusted age (change in slope 0.075, SE 0.029, P = .01) after participants were shown the risk message. There was no significant change (change in slope –0.027, SE 0.031, P = .38) for those participants in the non-HA risk message conditions. A similar pattern was found for the comparative risk question (Q2), with a statistically significant change in the regression slope against the risk-adjusted age for the HA risk message groups (change in slope 0.100, SE 0.030, P = .001), but no significant change for the non-HA risk message groups (change in slope –0.029, SE 0.030, P = .34).

XSL•FO
RenderX

Figure 4. Change in risk perceptions for Q1 and 2 split by Heart-Age level.



Intentions to Change

Generalized linear modeling showed that change in intention to reduce saturated fat intake at week 2 compared with week 1 was significantly influenced by condition ($c^2_3 = 18.8$, $P < .001$). Multiple comparisons between conditions (with a Bonferroni adjustment to allow for the six comparisons) showed that participants in the PT condition had a much higher intention than those allocated to the CG (odds ratio, 21.20; 95% CI, 2.6–172.4) or the HA risk message condition (odds ratio, 0.04; 95% CI, 0.0054–0.42).

There were no significant differences between the conditions for intention to take a cholesterol ($P = .38$) or blood pressure test within the next month ($P = .90$). There was a significant gender-by-group interaction ($c^2_3 = 13.6$, $P = .004$). Comparisons within the interaction effect (with a Bonferroni multiplicity adjustment) indicated that women who received the HA risk message were more motivated than the women in the CG to get their cholesterol tested within the next month (odds ratio, 2.46; 95% CI, 1.14–5.28). The same was true for women when the HA+PT condition was compared against the CG (odds ratio, 2.60; 95% CI, 1.18–5.76). There was no significant effect of condition on intention to test blood pressure ($c^2_3 = 0.8$, $P = .85$) and no interaction with gender.

Self-Efficacy

The generalized linear model showed that action self-efficacy measured at week 2 differed significantly between the conditions ($c^2_3 = 16.6$, $P < .001$). This was due to participants in the PT group being more confident than those in the CG (odds ratio, 3.06; 95% CI, 1.40–6.66). This difference was not statistically significant at week 5 ($c^2_3 = 7.1$, $P = .07$). Maintenance

XSL•FO
RenderX

self-efficacy measured at week 5 was not significantly different between the four conditions (P = .45).

Planning and Outcome Expectancies

At week 5, there was no significant difference in the "how" (P = .87) or "when" (P = .60) to reduce saturated fat intake between the four conditions. There were no significant effects of conditions for any of the outcome expectancy items.

Feedback on the Intervention

At week 2, there was a significant difference in perceived trustworthiness (c^2_3 = 8.9, P = .03), with those receiving the HA+PT reporting the intervention to be less trustworthy than those receiving the PT alone (mean 5.6 vs 5.9). There was also a difference between conditions for "informative" (c^2_3 = 14.3, P = .003) with HA+PT being perceived as less informative than the HA alone or the CG (mean 5.8 vs 6.1 vs 6.04). There was an overall difference in "worried" scores (c^2_3 = 4.8, P = .03). The HA+PT (mean 4.6) and the HA risk message participants (mean 4.7) were more worried than the PT participants (mean 4.0). All other feedback items were not significant. At week 5, there was still a significant difference between the conditions for "interesting" (c^2_3 = 8.6, P = .04), with the HA+PT participants still reporting the experience as less interesting than those receiving HA alone (mean 5.4 vs 5.7).

Discussion

Principal Results

In this study, a fully automated planning tool was successfully used by participants to form a set of health plans. The planning tool boosted self-efficacy and intention and reduced perceived saturated fat intake for one of the measures (TIS) but not the other (FFQ). A CVD risk message improved people's awareness of their risk relative to their age. Contrary to our expectations, combining a CVD risk message with the planning tool did not lead to bigger reductions in saturated fat intake than when they were presented on their own.

In line with theories of behavior change [36-38], the planning tool was better than the control group at increasing self-perceived consumption of low saturated fat foods (TIS). The same finding was not true for our second measure of saturated fat intake (FFQ). Also, participants in all conditions reported a change in percentage saturated fat intake measured by the FFQ, whereas participants in all conditions apart from the control group reported a change in their TIS score. Similar findings in terms of discrepancies between the FFQ and self-perceived items have been reported before [17]. This implies that the two self-perceived saturated fat intake items were better able than the FFQ measure to differentiate between the conditions. However, both come with limitations, which we discuss in the next section.

In line with our hypothesis, the planning tool was also better than the control group at boosting participants' intentions to reduce saturated fat [18] and action self-efficacy in the short term [19,20]. However, maintenance self-efficacy did not differ between the conditions at follow-up. This might be because participants who formed plans and encountered difficulties needed further support (eg, coping plans) to maintain their healthy eating. A previous study found that action plans are more effective at the early stages of change, while coping plans are instrumental at later stages [37].

In support of previous studies, receiving the heart-age risk message led to more appropriate risk perceptions [4,5,44], linked to participants' risk relative to their age group. Presentation of risk information also increased women's intentions to test their cholesterol. The latter finding is important because people who are aware of their cholesterol levels can receive more precise risk estimates.

Contrary to our expectations, combining the heart-age risk message with the planning tool (HA+PT) did not lead to a bigger reduction in saturated fat intake. A mismatch might have been created between the global CVD risk message and the specific target plan, confusing smokers with a high heart-age, who saw smoking cessation as the primary route to better health rather than diet. Alternatively, cognitive overload might have confounded the impact of HA+PT on saturated fat intake [45,46]. The length of time spent interacting online may also have been a factor, with the HA+PT taking the longest (12.47 minutes vs 10.91 for HA and 7.84 for PT). Future research could explore whether there is a benefit from reducing cognitive load through the use of a delay between presenting risk information and forming plans.

Limitations, Advantages, and Future Studies

The impact of conditions on our two measures of saturated fat intake changes was inconsistent, and this could be due to the limitations present in the FFQ and the TIS. Underreporting of food consumption is a recurrent challenge for FFQs and is most pronounced among overweight and obese people [47]. Also, FFQs were initially designed to estimate individual intake relative to a population rather than to detect small changes in individual dietary intake [33,48], for which they might not be sufficiently sensitive. The present FFQ did not account for individual variation in portion sizes but instead assumed the average portion of the UK population [33], which might differ from portions consumed by our obese participants.

On the other hand, self-perceived items like the TIS have been designed to detect differences between conditions in experimental studies [49]. However, some have claimed that reported changes are influenced by demand characteristics [50], with participants in more active conditions being more aware of study aims and so responding differently. Two previous studies counter the argument of demand characteristics by showing no difference between conditions for awareness of the study's hypothesis or feelings of obligation to comply [17,21]. Further research is needed to improve our ability to measure change in dietary intake (eg, through more objective measures).

To our knowledge, our planning tool is the first fully automated system to test online if–then plans in the format of an interactive volitional help sheet. An advantage of our approach was that participants could choose more personally relevant situations [30] from the list, promoting a sense of autonomy [51]. However, a disadvantage is that the list did not include highly

XSL•FO
RenderX

idiosyncratic situations that a participant might have entered through a free-text entry approach. Future studies could evaluate the relative impact of guiding participants to appropriate cues versus giving them complete autonomy.

As this was the first evaluation of a fully automated PT, we used a completers, per-protocol analysis, which, although limiting interpretation of the application of our results, allowed us to focus on the impact of the tool when used appropriately. Further research is needed to test the effectiveness of an implementation intention-based automated PT at a population level (via intention-to-treat analysis) over longer periods of time and to evaluate the impact of reminders [52-55].

Another advantage of the current study was that we assessed risk perceptions at two time points, giving us the opportunity to measure change in risk perceptions. Also, whereas previous research has used fictitious illnesses and hypothetical scenarios to communicate risk [23,41], our study risk corresponded to participants' personal characteristics, making it more relevant. To the best of our knowledge this is the first study to combine implementation intentions with personally relevant health risk information through a Web-based medium.

Conclusions

Web-based tools provide a good opportunity to present risk information and plan behavior change. In the present study, the HA risk message helped improve obese people's awareness of risk relative to their age, and the PT reduced levels of perceived saturated fat intake. Future research is required to identify ways of matching motivational and volitional strategies to change behavior.

Acknowledgments

Andrew Hinton and Steve Baker from Tessella PLC had full access to all of the data and take responsibility for the integrity of the data. Virtual Surveys PLC was in charge of the online recruitment of participants.

Conflicts of Interest

This exploratory study was funded by Unilever, in which several authors are employees. Heart-Age is a commercially available product developed by Unilever and used as the basis of risk communication, while the planning tool is a non-commercially available research vehicle.

Multimedia Appendix 1

Research protocol.

[PDF File (Adobe PDF File), 158KB - jmir_v13i4e100_app1.pdf]

Multimedia Appendix 2

Ethics committee approval document.

[PDF File (Adobe PDF File), 158KB - jmir_v13i4e100_app2.pdf]

References

1. US Department of Health and Human Services. Health, United States 2003: With Chartbook on Trends in the Health of Americans (Health United States). Washington, DC: Congress; 2003.
2. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [Medline: 6392204]
3. Schwarzer R. Modelling health behaviour change: how to predict and modify the adoption and maintenance of health behaviours. Appl Psychol Int Rev 2008;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
4. Soureti A, Hurling R, Murray P, van Mechelen W, Cobain M. Evaluation of a cardiovascular disease risk assessment tool for the promotion of healthier lifestyles. Eur J Cardiovasc Prev Rehabil 2010 Oct;17(5):519-523. [doi: 10.1097/HJR.0b013e328337ccd3] [Medline: 20195154]
5. Parkes G, Greenhalgh T, Griffin M, Dent R. Effect on smoking quit rate of telling patients their lung age: the Step2quit randomised controlled trial. BMJ 2008 Mar 15;336(7644):598-600 [FREE Full text] [doi: 10.1136/bmj.39503.582396.25] [Medline: 18326503]
6. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
7. Milne S, Orbell S, Sheeran P. Combining motivational and volitional interventions to promote exercise participation: protection motivation theory and implementation intentions. Br J Health Psychol 2002 May;7(Pt 2):163-184. [doi: 10.1348/135910702169420] [Medline: 14596707]
8. Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement: a meta-analysis of effects and processes. Adv Exp Soc Psychol 2006;38:69-111. [doi: 10.1016/S0065-2601(06)38002-1]

XSL•FO
RenderX

9. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 2000 Oct;27(5):591-615. [Medline: 11009129]
10. Orbell S, Sheeran P. Motivational and volitional processes in action initiation: a field study of the role of implementation intentions. J Appl Soc Psychol 2000;30(4):780-797. [doi: 10.1111/j.1559-1816.2000.tb02823.x]
11. Webb TL, Sheeran P. Mechanisms of implementation intention effects: the role of goal intentions, self-efficacy, and accessibility of plan components. Br J Soc Psychol 2008 Sep;47(Pt 3):373-395. [doi: 10.1348/014466607X267010] [Medline: 18096108]
12. Webb TL, Sheeran P. How do implementation intentions promote goal attainment? A test of component processes. J Exp Soc Psychol 2007;43(2):295-302. [doi: 10.1016/j.jesp.2006.02.001]
13. Gollwitzer PM. Goal achievement: the role of intentions. Eur Rev Soc Psychol 1993;4(1):141-185. [doi: 10.1080/14792779343000059]
14. Aarts H, Dijksterhuis A, Midden C. To plan or not to plan? Goal achievement or interrupting the performance of mundane behaviours. Eur J Social Psychol 1999;29(8):971-979. [doi: 10.1002/(SICI)1099-0992]
15. Aarts H, Dijksterhuis A. Habits as knowledge structures: automaticity in goal-directed behavior. J Pers Soc Psychol 2000 Jan;78(1):53-63. [Medline: 10653505]
16. Oettingen G, Honig G, Gollwitzer PM. Effective self-regulation of goal attainment. Int J Educ Res 2000;33(7-8):705-732. [doi: 10.1016/S0883-0355(00)00046-X]
17. Chapman J, Armitage CJ, Norman P. Comparing implementation intention interventions in relation to young adults' intake of fruit and vegetables. Psychol Health 2009 Mar;24(3):317-332. [doi: 10.1080/08870440701864538] [Medline: 20204996]
18. Armitage CJ. Evidence that implementation intentions reduce dietary fat intake: a randomized trial. Health Psychol 2004 May;23(3):319-323. [doi: 10.1037/0278-6133.23.3.319] [Medline: 15099174]
19. Murray TC, Rodgers WM, Shawn NF. Examining implementation intentions in an exercise intervention: the effects on adherence and self-efficacy in a naturalistic setting. J Appl Soc Psychol 2009;39:2303-2320. [doi: 10.1111/j.1559-1816.2009.00527.x]
20. Rodgers WM, Hall CR, Blanchard CM, McAuley E, Munroe KJ. Task and scheduling self-efficacy as predictors of exercise behavior. Psychol Health 2002;17(4):405-416. [doi: 10.1080/0887044022000004902]
21. Budden JS, Sagarin BJ. Implementation intentions, occupational stress, and the exercise intention-behavior relationship. J Occup Health Psychol 2007 Oct;12(4):391-401. [doi: 10.1037/1076-8998.12.4.391] [Medline: 17953497]
22. Soureti A, Murray P, Cobain M, Chinapaw M, van Mechelen W, Hurling R. A study exploring the impact of on an online planning tool and mobile text reminder service on healthier dietary behaviours. J Med Internet Res 2011. [doi: 10.2196/jmir.1773]
23. Prestwich A, Ayres K, Lawton R. Crossing two types of implementation intentions with a protection motivation intervention for the reduction of saturated fat intake: a randomized trial. Soc Sci Med 2008 Nov;67(10):1550-1558. [doi: 10.1016/j.socscimed.2008.07.019] [Medline: 18771834]
24. Renner B, Knoll N, Schwarzer R. Age and body make a difference in optimistic health beliefs and nutrition behaviours. Int J Behav Med 2000;7(2):143-159. [doi: 10.1207/S15327558IJBM0702_4]
25. Varnhagen CK, Gushta M, Daniels J, Peters TC, Parmar N, Law D, et al. How informed is online informed consent? Ethics Behav 2005;15(1):37-48. [doi: 10.1207/s15327019eb1501_3] [Medline: 16127857]
26. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008 Feb 12;117(6):743-753 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.699579] [Medline: 18212285]
27. Armitage CJ. A volitional help sheet to encourage smoking cessation: a randomized exploratory trial. Health Psychol 2008 Sep;27(5):557-566. [doi: 10.1037/0278-6133.27.5.557] [Medline: 18823182]
28. Prochaska JO. Systems of Psychotherapy: A Transtheoretical Analysis. Homewood, IL: Dorsey Press; 1984.
29. UK Food Standards Agency. Saturated fat URL: http://www.eatwell.gov.uk/healthydiet/fss/fats/satfat/ [accessed 2010-02-01] [WebCite Cache ID 5nDvA2JkB]
30. Adriaanse MA, de Ridder DT, de Wit JB. Finding the critical cue: implementation intentions to change one's diet work best when tailored to personally relevant reasons for unhealthy eating. Pers Soc Psychol Bull 2009 Jan;35(1):60-71. [doi: 10.1177/0146167208325612] [Medline: 19106078]
31. Rossi SR, Greene GW, Rossi JS, Plummer BA, Benisovich SV, Keller S, et al. Validation of decisional balance and situational temptations measures for dietary fat reduction in a large school-based population of adolescents. Eat Behav 2001;2(1):1-18. [Medline: 15001046]
32. Velicer WF, Diclemente CC, Rossi JS, Prochaska JO. Relapse situations and self-efficacy: an integrative model. Addict Behav 1990;15(3):271-283. [Medline: 2378287]
33. Margetts BM, Cade JE, Osmond C. Comparison of a food frequency questionnaire with a diet record. Int J Epidemiol 1989 Dec;18(4):868-873. [Medline: 2621023]
34. Armitage CJ, Conner M. The theory of planned behaviour: assessment of predictive validity and 'perceived control'. Br J Soc Psychol 1999;38(1):35-54. [doi: 10.1348/014466699164022]

XSL•FO
RenderX

35. Thompson RL, Margetts BM. Comparison of a food frequency questionnaire with a 10-day weighed record in cigarette smokers. Int J Epidemiol 1993 Oct;22(5):824-833. [Medline: 8282461]
36. Renner B, Schwarzer R. International University Bremen and Freie Universität Berlin. Risk and Health Behaviors: Documentation of the Scales of the Research Project: "Risk Appraisal Consequences in Korea" (RACK). 2nd edition URL: http://www.gesundheitsrisiko.de/docs/RACKEnglish.pdf [accessed 2011-06-27] [WebCite Cache ID 5zkw6NCFj]
37. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention-behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005;20(2):143-160. [doi: 10.1080/08870440512331317670]
38. Schwarzer R, Sniehotta FF, Lippke S, Luszczynska A, Scholz U, Schüz B, et al. Freie Universität Berlin. 2003 Nov 1. On the Assessment and Analysis of Variables in the Health Action Process Approach: Conducting an Investigation URL: http://web.fu-berlin.de/gesund/hapa_web.pdf [accessed 2011-06-27] [WebCite Cache ID 5zkvk39uc]
39. De Bourdeaudhuij I, Brug J. Tailoring dietary feedback to reduce fat intake: an intervention at the family level. Health Educ Res 2000 Aug;15(4):449-462 [FREE Full text] [Medline: 11066462]
40. Brug J, Glanz K, Van Assema P, Kok G, van Breukelen GJ. The impact of computer-tailored feedback and iterative feedback on fat, fruit, and vegetable intake. Health Educ Behav 1998 Aug;25(4):517-531. [Medline: 9690108]
41. Fair AK, Murray PG, Thomas A, Cobain MR. Using hypothetical data to assess the effect of numerical format and context on the perception of coronary heart disease risk. Am J Health Promot 2008;22(4):291-296. [doi: 10.4278/061030140R2.1] [Medline: 18421894]
42. Anonymous. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2000 Dec 20;284(23):3043-3045. [Medline: 11122593]
43. BBC Health. 2010 Jan. Nutrition: Healthy Eating: Fats and Sugars URL: http://www.bbc.co.uk/health/healthy_living/nutrition/healthy_fatsugar.shtml [accessed 2010-03-11] [WebCite Cache ID 5o9XPu0rQ]
44. Grover SA, Lowensteyn I, Joseph L, Kaouache M, Marchand S, Coupal L, Cardiovascular Health Evaluation to Improve Compliance and Knowledge Among Uninformed Patients (CHECK-UP) Study Group. Patient knowledge of coronary risk profile improves the effectiveness of dyslipidemia therapy: the CHECK-UP study: a randomized controlled trial. Arch Intern Med 2007 Nov 26;167(21):2296-2303 [FREE Full text] [doi: 10.1001/archinte.167.21.2296] [Medline: 18039987]
45. Dewitte S, Pandalaere M, Briers B, Warlog L. Social Science Research Network. 2005 Oct. Cognitive Load Has Negative After Effects on Consumer Decision Making URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=813684 [accessed 2011-10-31] [WebCite Cache ID 62qh2PBur]
46. Ward A, Mann T. Don't mind if I do: disinhibited eating under cognitive load. J Pers Soc Psychol 2000 Apr;78(4):753-763. [Medline: 10794378]
47. Garriguet D. Under-reporting of energy intake in the Canadian Community Health Survey. Health Rep 2008 Dec;19(4):37-45 [FREE Full text] [Medline: 19226926]
48. Cade JE, Burley VJ, Warm DL, Thompson RL, Margetts BM. Food-frequency questionnaires: a review of their design, validation and utilisation. Nutr Res Rev 2004 Jun;17(1):5-22. [doi: 10.1079/NRR200370] [Medline: 19079912]
49. Chapman J, Armitage CJ. Evidence that boosters augment the long-term impact of implementation intentions on fruit and vegetable intake. Psychol Health 2010 Mar;25(3):365-381. [doi: 10.1080/08870440802642148] [Medline: 20204966]
50. Faith MS, Wong FY, Allison DB. Demand characteristics of the research setting can influence indexes of negative affect-induced eating in obese individuals. Obes Res 1998 Mar;6(2):134-136. [Medline: 9545020]
51. Koestner R, Horberg EJ, Gaudreau P, Powers T, Di Dio P, Bryan C, et al. Bolstering implementation plans for the long haul: the benefits of simultaneously boosting self-concordance or self-efficacy. Pers Soc Psychol Bull 2006 Nov;32(11):1547-1558. [doi: 10.1177/0146167206291782] [Medline: 17030894]
52. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]
53. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007 Apr;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
54. Ware LJ, Hurling R, Bataveljic O, Fairley BW, Hurst TL, Murray P, et al. Rates and determinants of uptake and use of an internet physical activity and weight management program in office and manufacturing work sites in England: cohort study. J Med Internet Res 2008;10(4):e56 [FREE Full text] [doi: 10.2196/jmir.1108] [Medline: 19117828]
55. Prestwich A, Perugini M, Hurling R. Can implementation intentions and text messages promote brisk walking? A randomized trial. Health Psychol 2010 Jan;29(1):40-49. [doi: 10.1037/a0016993] [Medline: 20063934]

Abbreviations

CG: control group
CVD: cardiovascular disease
FFQ: food frequency questionnaire

XSL•FO
RenderX

HA: Heart-Age
PT: planning tool
Q1: individual risk question
Q2: comparative risk question
TIS: 2-item scale

Edited by G Eysenbach; submitted 01.06.10; peer-reviewed by G Norman, K Funk, J Van 't Riet; comments to author 23.06.10; revised version received 08.09.11; accepted 25.09.11; published 24.11.11.

Please cite as:
Soureti A, Murray P, Cobain M, van Mechelen W, Hurling R
Web-Based Risk Communication and Planning in an Obese Population: Exploratory Study
J Med Internet Res 2011;13(4):e100
URL: http://www.jmir.org/2011/4/e100/
doi:10.2196/jmir.1579
PMID:22126827

©Anastasia Soureti, Peter Murray, Mark Cobain, Willem van Mechelen, Robert Hurling. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effects of eHealth Interventions on Medication Adherence: A Systematic Review of the Literature

Annemiek J Linn[1], MSc; Marcia Vervloet[2], MSc; Liset van Dijk[2], PhD; Edith G Smit[1], PhD; Julia CM Van Weert[1], PhD

[1]Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

[2]Netherlands Institute for Health Services Research, Utrecht, Netherlands

Corresponding Author:
Marcia Vervloet, MSc
Netherlands Institute for Health Services Research
P.O. Box 1568
Utrecht, 3500 BN
Netherlands
Phone: 31 30 2729713
Fax: 31 30 2729729
Email: M.Vervloet@nivel.nl

Abstract

Background: Since medication nonadherence is considered to be an important health risk, numerous interventions to improve adherence have been developed. During the past decade, the use of Internet-based interventions to improve medication adherence has increased rapidly. Internet interventions have the potential advantage of tailoring the interventions to the needs and situation of the patient.

Objective: The main aim of this systematic review was to investigate which tailored Internet interventions are effective in improving medication adherence.

Methods: We undertook comprehensive literature searches in PubMed, PsycINFO, EMBASE, CINAHL, and Communication Abstracts, following the guidelines of the Cochrane Collaboration. The methodological quality of the randomized controlled trials and clinical controlled trials and methods for measuring adherence were independently reviewed by two researchers.

Results: A total of 13 studies met the inclusion criteria. All included Internet interventions clearly used moderately or highly sophisticated computer-tailored methods. Data synthesis revealed that there is evidence for the effectiveness of Internet interventions in improving medication adherence: 5 studies (3 high-quality studies and 2 low-quality studies) showed a significant effect on adherence; 6 other studies (4 high-quality studies and 2 low-quality studies) reported a moderate effect on adherence; and 2 studies (1 high-quality study and 1 low-quality study) showed no effect on patients' adherence. However, most studies used self-reported measurements to assess adherence, which is generally perceived as a low-quality measurement. In addition, we did not find a clear relationship between the quality of the studies or the level of sophistication of message tailoring and the effectiveness of the intervention. This might be explained by the great difference in study designs and the way of measuring adherence, which makes results difficult to compare. There was also large variation in the measured interval between baseline and follow-up measurements.

Conclusion: This review shows promising results on the effectiveness of Internet interventions to enhance patients' adherence to prescribed long-term medications. Although there is evidence according to the data synthesis, the results must be interpreted with caution due to low-quality adherence measurements. Future studies using high-quality measurements to assess medication adherence are recommended to establish more robust evidence for the effectiveness of eHealth interventions on medication adherence.

(J Med Internet Res 2011;13(4):e103) doi:10.2196/jmir.1738

KEYWORDS

Internet interventions; medication adherence; compliance; systematic review; tailoring; eHealth; effects; RCT

XSL•FO
RenderX

Introduction

Recent reports of the World Health Organization and the National Institute for Health and Clinical Excellence reveal that 30%–50% of patients with chronic illnesses do not adhere to prescribed medication [1]. Other studies also show that rates of nonadherence are very high and depend on the type of disease. The highest adherence rates are found for patients with human immunodeficiency virus infection, while diabetes patients have the lowest rate [2]. As such, medication nonadherence can be considered an important health care problem. This is especially true for patients with a chronic illness because medication adherence is a crucial factor in the effectiveness of a therapy [2]. Consequently, many patient-centered interventions are developed to improve adherence, and the impact of the Internet in the development of these interventions is increasing. It is therefore important to understand how these interventions work and to know whether they are effective in improving adherence. To our knowledge, no recent review has studied the effectiveness of patient-centered Internet interventions on patients' medication adherence. Therefore, we conducted a systematic literature study in which we reviewed evidence from studies on Internet interventions that were developed to assist patients in their medication management. The purpose of our study was fourfold: first, to gain insight into the current stage of development of these interventions; second, to assess the included studies for their effectiveness on medication adherence; third, to investigate to what degree adherence is determined by the characteristics of the intervention; and fourth, to investigate whether there is a relationship between the characteristics of the study and the reported effectiveness of the interventions.

Different terms are used in the literature to describe the concept of adherence—for example, compliance, adherence, and persistence. They have all been used to indicate that the patient is using the medication following the prescribed regimen. These terms differ in exact meaning. In this paper, we use the term adherence. *Adherence* is defined as the extent to which the patient's behavior matches the agreed recommendations of the prescriber [3]. According to this definition, nonadherence is a wide concept that varies from missing an occasional dose to never taking the prescribed medications [3]. Patients have different reasons for being nonadherent. These different reasons have something in common: the patient does not execute the treatment plan and does not persist. Execution is a continuous process where the actual dosing history corresponds to the ideal doses [4,5].

To improve adherence and develop target interventions, it is important to address the specific reasons why a patient is not able or willing to execute the treatment plan. From this perspective, interventions should be personalized or tailored to address individual needs and beliefs. The definition of tailoring describes the features that make tailored health messages different from other approaches: "It is assessment-based and as a result the message can be individual-focused" [6]. In other words, tailoring is based on gathering and assessing personal data related to health outcomes or several determinants in order to determine the most effective strategy to meet that person's needs [6]. With these characteristics, a tailored message is able to provide personal feedback, commands greater attention, is processed more deeply, and is perceived as more likable by patients than a general message [7,8]. Because of these possibilities, tailored health messages are also more likely than generic information to be read, remembered, and viewed as personally relevant [6,9].

Computer technologies can be used to tailor health messages to the personal situation of the patient and might therefore contribute significantly to the development of tailored message strategies. The Internet is potentially a powerful medium for delivering those tailored messages. The management of a chronic disease should be personalized to an individual because the person is ultimately responsible for the success of the intervention [10]. The technology provides an opportunity to tailor the information in several formats and modalities, which enhances the user's experience of the material and will result in a better understanding [7,8]. Moreover, Internet interventions have the advantage that they can provide interactive and responsive programs [10]. These interventions can provide effective data and information provision and retrieval. The advantages of tailored message strategies can contribute to the incorporation of interactive and continued self-monitoring, feedback, and information exchange, which play an increasingly important role in changing patients' behavior.

Methods

For this review, we used the guidelines of the Cochrane Collaboration to assess the studies on their internal validity and to summarize the existing evidence about Internet interventions to improve medication adherence in patients. The Cochrane Collaboration method is described in more detail in the Cochrane Handbook for Systematic Reviews of Interventions [11].

Inclusion Criteria

We included a study when the following inclusion criteria were met: (1) the study described a patient-centered Internet intervention, (2) the study described an intervention for patients who use prescribed medication for a chronic condition, (3) at least one of the outcome measures was adherence, (4) the study was quantitative, and (5) the study was published in either the English or Dutch language.

Search Methods

We conducted a systematic literature search to identify articles containing information about the effect of Internet interventions on medication adherence. Comprehensive literature searches were undertaken in the databases PubMed, PsycINFO, EMBASE, CINAHL, and Communication Abstracts.

The search strategies used the following keywords: (medication therapy management OR medication adherence OR patient compliance OR self-care) AND (internet) AND (intervention study OR randomized controlled trial OR clinical controlled trial).

We then continued with the snowball method by looking for references in publications, especially those of the included studies and reviews on interventions to promote adherence. The

XSL•FO
RenderX

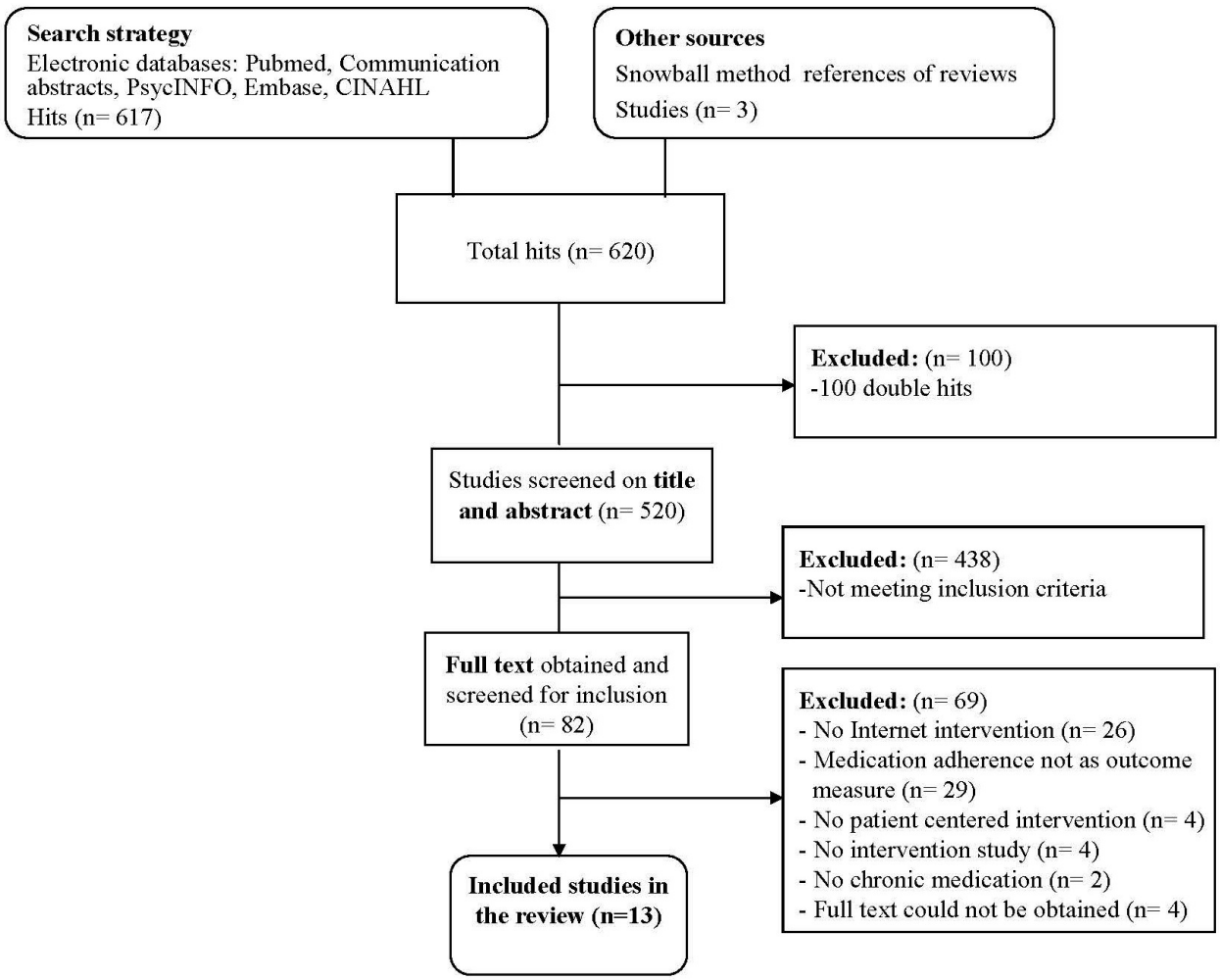
search was conducted in June 2010. Since Internet intervention is a relatively new topic, no time limits were applied.

Application of the search strategy to the specified databases resulted in a total of 620 hits (Table 1). In total, we selected 13 studies from these results.

Table 1. Results of database searches.

Source	Hits per strategy	Unique studies	Relevant studies
PubMed	388	388	11
Communication Abstracts	0	0	0
PsycINFO	47	40	0
EMBASE	169	82	0
Snowball method	3	0	1
CINAHL	13	10	1
Total	620	520	13

Figure 1. Flow diagram of study search and selection.



XSL•FO
RenderX

Reference Manager version 11.0 (Thomson Reuters, New York, NY, USA) was used to manage the citations. Duplications were logged, leaving 520 unique results (see Figure 1). On the basis of title and abstract, two researchers (pairs of AL, MV, LvD, JvW) independently selected studies for inclusion. If the study seemed to meet the inclusion criteria or if there were doubts about the inclusion, the full text of the article was obtained. Based on the full articles two reviewers independently reviewed whether these studies fit all the inclusion criteria. Disagreements were solved by discussions between the two researchers. For a more detailed description of the excluded studies see Multimedia Appendix 1.

Assessment of Methodological Quality

The methodological quality of included randomized controlled trials (RCTs) and clinical controlled trials (CCTs) was independently reviewed by two researchers (AL and JvW) using the list from the Cochrane Collaboration Back Review Group [12]. The list consists of 11 criteria for internal validity, namely:

- 3 criteria regarding selection bias: whether (a) randomization was adequate, (b) treatment allocation was concealed, and (c) groups were similar at baseline regarding the most important indicators,
- 4 criteria for performance bias: whether (d) patients were blinded to the intervention, (e) care provider was blinded to the intervention, (g) co-interventions were avoided, and (h) compliance with the intervention was acceptable,
- 2 criteria regarding attrition bias: whether (i) the dropout rate after baseline was acceptable, and (k) the analysis included an intention-to-treat analysis, and
- 2 criteria for detection bias: whether (f) the outcome assessor was blinded to the intervention, and (j) outcome assessments in all groups were similar.

For each included study, all criteria were scored as “yes,” “no,” or “unclear.” All unclear scores were later rated as “no.” Studies were rated as high quality (HQ) when at least 6 of the 11 criteria for internal validity were met. Otherwise, studies were considered of low quality (LQ). Disagreements were discussed until consensus was reached. If disagreement or indistinctness persisted a third reviewer (LvD) was consulted.

In addition, two researchers (AL and LvD) independently assessed the quality of the methods for measuring adherence to a medical regimen. A standard method to assess adherence does not exist and every method has its limits [13,14]. In clinical trials, adherence can be measured based on, for example, interviews, diary, questionnaire-based self-reporting, prescription refills, pill counts, or electronic monitoring [14,15]. We categorized the measurements in high- and low-quality adherence assessments based on previous findings concerning the objectivity of these measurements [13-15]. In this review electronic monitoring and physiological/biomedical measures are defined as high-quality adherence assessment. These measurements are considered the most objective standard [15]. Previous research has shown that data from pill counts and electronic monitoring are strongly correlated [16]. Yet others consider pill counts not to be accurate [17,18]. In addition, meta-analyses have shown that self-reported adherence is also strongly correlated with electronic monitoring [19,20]. Like pill counts, the accuracy of self-reported measurements is debatable. Some argue that self-reports may be an accurate measurement for measuring adherence [21,22], while others state that the use of self-reported measurements is not an accurate method [14,15,23,24]. Taking all arguments into account, we considered self-measurements, such as questionnaires, pill counts, prescription refills, interviews, and diaries, to be most subjective for measuring adherence [13]. We therefore considered these measurements low-quality adherence assessment. However, if two or more different low-quality adherence measurements were used in the same study, such as a combination of questionnaires and prescription refills, the method was considered high-quality adherence assessment.

Data Extraction

One researcher (AL) documented the following characteristics of the included studies: (1) method (type of study), (2) participants (total number of participants, sex per group, mean age per group, type of disease), (3) intervention (name of experimental intervention, name of control condition, period, number of times/minutes per week), (4) outcome measures (type of outcome measures, time of measurement), (5) results (short description), and (6) author’s conclusion.

Data Synthesis

Due to diversity in the features of the interventions and the methods used to measure adherence, it was not possible to pool the data. Therefore, we conducted a best-evidence synthesis (see Textbox 1) based on [12] and adapted by a Dutch study [25].

The best-evidence synthesis was conducted by attributing various levels of evidence to the effectiveness of the interventions. The synthesis takes into account the design, methodological quality, and outcomes of the studies. Textbox 1 shows that at least 1 HQ RCT or 2 HQ CCTs were needed to establish robust evidence for the effectiveness of an intervention.

XSL•FO
RenderX

Textbox 1. Principles of Best-Evidence Synthesis.

Evidence:

Provided by consistent, statistically significant findings in outcome measures in at least 2 high-quality randomized controlled trials (RCTs).

Moderate evidence:

Provided by consistent, statistically significant findings in outcome measures in at least 1 high-quality RCT and at least 1 low-quality RCT or high-quality CCT.

Limited evidence:

Provided by statistically significant findings in outcome measures in at least 1 high-quality RCT

Or

Provided by consistent, statistically significant findings in outcome measures in at least 2 high-quality CCTs (in the absence of high-quality RCTs).

Indicative findings:

Provided by statistically significant findings in outcome measures in at least 1 high-quality CCT or low-quality RCT (in the absence of high-quality RCTs).

No/insufficient evidence:

If the number of studies that have significant findings is <50% of the total number of studies found within the same category of methodological quality and study design

Or

In case the results of eligible studies do not meet the criteria for one of the above stated levels of evidence

Or

In case of conflicting (statistically significantly positive and statistically significantly negative) results between RCTs and CCTs

Or

In case of no eligible studies.

Sensitivity Analysis

We conducted a sensitivity analysis to indentify how sensitive the results of the best-evidence syntheses were to changes in the way the study quality was assessed. For the sensitivity analysis, the best-evidence synthesis was repeated in two different ways, using the following principles: (1) LQ studies were excluded, (2) studies were rated as HQ if they met at least 4 of the 11 criteria of internal validity instead of 6.

We then compared the results of the sensitivity analysis with the results of the best-evidence synthesis and described the sensitivity of the results [25,26].

Effectiveness

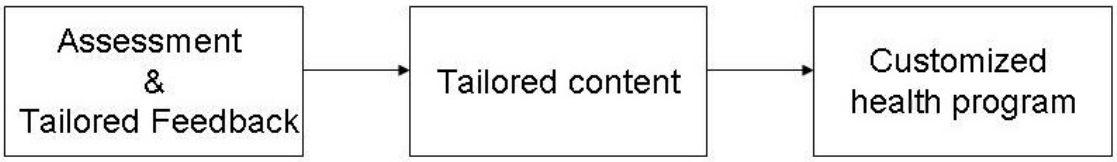
Study effectiveness was categorized as significant effect on adherence, moderate effect on adherence, and no effect on adherence. We defined a study effect as moderate if the authors reported a positive effect of the intervention on adherence but there were limitations, such as the following: improvement of adherence was found only in a subgroup of the intervention group; adherence was measured indirectly (eg, the study drew conclusions about the use of beta-agonist indicating that adherence was improved); or the significance of the results to medication adherence was not tested, but the authors used convincing arguments to explain the effectiveness of the intervention (see results section for explanation per study).

Intervention: Tailoring Level of Sophistication of the Website

Tailored Internet interventions differ in how they deliver their message [7]. The difference is based on the sophistication of the way the message is tailored. We categorized the interventions in being low, moderate, or high in sophistication. Some interventions involve a form of online assessments (low sophistication), and others use online assessments, tailored feedback, and content matching (moderate sophistication). The third group of interventions provides instant feedback and a complex tailored health program with several tools and activities that would enable patients to achieve their health goals (high sophistication) (see Figure 2) [7].

XSL•FO
RenderX

Figure 2. Continuum level of sophistication of tailored intervention.



Results

The main characteristics of the included studies are presented in Table 2 [27-39] and further described below (for a more detailed description of the included studies see Multimedia Appendix 2).

XSL•FO
RenderX

Table 2. Characteristics of included studies.

Study; method	Intervention[a]	Participants; sex; mean age	Adherence measurement; timing of measuring adherence	Main conclusion
Artinian [27]; RCT[b]	Web-based monitoring system; tailored content; nature of expert/therapist contact	N = 18 (17 males; mean age 68 years); intervention group n = 9, control group n = 9	Pill counts; baseline and 3 months	Medication compliance rate was 94% for the monitor group as measured by the monitor system
Jan [28]; RCT[b]	Blue Angel for Asthma Kids variability; tailored content; nature of expert/therapist contact	N = 164, intervention group n = 88 (35 males; mean age 10.9 years); control group n = 76 (28 males; mean age 9.9 years)	Self-reported at baseline and 12 weeks	The Blue Angel for Asthma Kids has the potential for improving asthma outcome compared with conventional treatment over a period of 12 weeks
Chan [29]; RCT[b]	Customized educational and monitoring Web site; tailored content; nature of expert/therapist contact	N = 120; intervention group n = 60 (37 males; mean age 10.2 years); control group n = 60 (38 males; mean age 9.0 years)	Computerized prescription refill record at baseline, 26 weeks, and 52 weeks	No difference in adherence between groups
Chan [30]; RCT[b]	Customized educational and monitoring Web site; tailored content; nature of expert/therapist contact	N = 10; intervention group n = 5 (1 male; mean age 6.6 years); control group n = 5 (4 males; mean age 8.7 years)	Self-reported asthma diary and computerized prescription refill record at 90 days and 180 days	After the intervention, the use of beta-agonist decreased, which is an indication of better adherence
Joseph [31]; RCT[b]	Web-based asthma management program; tailored content; user control	N = 314 (36.6% male; mean age 15.3 years); intervention group n = 162; control group n = 52	Self-reported at baseline and 12 months	Positive changes in controller medication adherence were seen
Ross [32]; RCT[b]	SPPARO (System Providing Access to Records Online); tailored content; nature of expert/therapist contact	N = 104; intervention group n = 54 (80% male; mean age 57 years); control group n = 50 (74% male; mean age 55 years)	Self-reported at baseline, 6 months, and 12 months	Providing patients access to an online medical record improved adherence
Cherry [33]; prospective design	Telemedicine diabetes disease management program; tailored content; nature of expert/therapist contact	Intervention group n = 169 (39 males; mean age 53 years); historical group (usual care)	Self-reported	Outcomes offer encouraging evidence that telemedicine technology coupled with daily remote monitoring may improve appropriate use of health care services
Guendelman [34]; RCT[b]	Health Buddy, an interactive device connected to a home telephone; tailored content; nature of expert/therapist contact	N = 134; intervention group (40 males; mean age 12.2 years); control group (37 males; mean age 12.0 years)	Self-reported at baselines, 6 weeks, and 12 weeks	Patients were more likely to take their asthma medication without additional reminders
DeVito Dabbs [35]; RCT[b]	Pocket Personal Assistant for Tracking Health (PATH); tailored content; nature of expert/therapist contact	N = 30; intervention group n = 15 (60% male; mean age 55 years); control group n = 15 (60% male; mean age 57 years)	Self-reported at baseline and 2 months	Patients who received the PATH were more likely to show high adherence to the medical regimen
Van der Meer [36]; RCT[b]	Internet-based self-management program; customized health program; user control	N = 200; intervention group n = 101 (29% male; mean age 36 years); control group n = 99 (29% male; mean age 37 years)	Self-reported at baseline, 3 months, and 6 months	After 3 months asthma control improved
Van der Meer [37]; RCT[b]	Internet-based self-management program; customized health program; user control	N = 200; intervention group n = 111 (28 males; mean age 36 years); control group n = 89 (28 males; mean age 36.6 years)	Self-reported at baseline, 3 months, and 1 year	Weekly self-monitoring leads to improved asthma control in patients with partly and uncontrolled asthma at baseline and tailors asthma medication to individual patients' needs

XSL•FO
RenderX

Study; method	Intervention[a]	Participants; sex; mean age	Adherence measurement; timing of measuring adherence	Main conclusion
Dilorio [38]; Survey	WebEase; customized health program; user control	N = 35 (40% male; mean age 37.5 years)	Self-reported at baseline and 6 weeks	Participants showed some improvement in adherence following the program
Dew [39]; prospective design	Website including skills workshops, discussion group, ask an expert, question and answer, health tips, recourses, and references; customized health program; nature of expert/therapist contact	N = 64; intervention group n = 24 (18 males; mean age 45.8 years); control group n = 40 (30 males; mean age 57.5 years)	Self-reported at baseline and 4 months	The intervention appeared to be weakly associated with medical compliance change

[a] Sophistication of tailoring classification based on Figure 2.

[b] Randomized controlled trial.

Methodological Quality: Assessment of Internal Validity

For this review 10 RCTs were included, and 9 of them were assessed on their internal validity (Table 3 [27-39]). We included 1 RCT with no data on our primary outcome variable (ie, medication adherence) for the control group [27]. This means that for this tenth study, we could not assess validity criteria. Moreover, we reviewed 2 prospective cohort designs and 1 survey. A total of 7 RCTs met 6 or more of the 11 validity criteria and therefore qualified as HQ studies.

Table 3. Results of methodological quality.

Study	Validity criteria[a] met	Study quality[b]	Quality measurement adherence
Randomized clinical trials			
Artinian [27]	Not applicable[c]	Low	Low
Jan [28]	a, b, c, d, i, j	High	Low
Chan [29]	a, b, f, i, j	Low	Low
Chan [30]	a, b, c, f, h[d], i, j	High	High
Joseph [31]	a, b, c, d, e, h, i, j	High	Low
Ross [32]	a, b, c, d, e, h[e], i, j, k	High	Low
Guendelman [34]	a, b, d, i, j	Low	Low
DeVito Dabbs [35]	a, b, c, e, i, j	High	Low
Van der Meer [36]	a, b, c, i, j, k	High	Low
Van der Meer [37]	a, c, h[f], i, j, k	High	Low
Prospective design/clinical trial or cohort design			
Cherry [33]		Low	Low
Dew [39]		Low	Low
Survey			
Dilorio [38]		Low	Low

[a] a: randomization adequate; b: treatment allocation concealed; c: groups similar at baseline regarding most important indicators; d: patients blinded to intervention; e: care provider blinded to intervention; f: outcome assessor blinded to intervention; g: co-interventions avoided; h: compliance with intervention acceptable; i: dropout rate after baseline acceptable; j: outcome assessed similarly in all groups; k: intention-to-treat analysis included.

[b] That is, 6 of 11 validity criteria were met.

[c] No data on medication adherence for the control group and therefore judged as low quality.

[d] Compliance was acceptable in the first interval (<90 days).

[e] Compliance was acceptable in the first interval (6 months).

[f] Compliance was acceptable in the first interval (3 months).

XSL•FO
RenderX

Intervention: Tailoring Level of Sophistication of the Website

All Internet interventions reported computer-tailoring methods. Interventions were categorized as having low sophistication (online assessments), moderate sophistication (online assessments, tailored feedback, and content matching), and high sophistication (a more complex tailored health program) (see Figure 2) [7].

Online Assessment and Feedback

Online assessment and feedback are used in interventions with single-incident computer-assisted risk or health assessments. For example, feedback is emailed to the patients or provided online [7]. In addition, these interventions are brief and usually done once at the beginning of the intervention. None of the reviewed studies used online assessment and feedback.

Tailored Content

With tailored content a program provides (1) tailored text messages composed in a unique way according to how patients respond to certain questions, or (2) restricted access to content sections per patient [7]. Tailored content was used by 9 of the studies we reviewed; 1 study [28], Blue Angel for Asthma Kids, conducted an Internet-based interactive asthma educational and monitoring program in which patients were able to complete an electronic diary, record symptoms and need for rescue medication, and upload their videos when they were using their inhaler. Based on these outcomes, the program comprised both an action plan with a warning system and a written treatment plan. A similar customized educational and monitoring Website for patients with asthma was conducted in 2 studies [29,30], and 1 study [31] tested the asthma management program Puff City. The program used tailoring to alter behavior through individualized health messages based on the patients' beliefs, attitudes, and personal barriers to change or maintain the behavior. Another form of a tailored website, System Providing Access to Records Online (SPPARO), was examined in 1 study [32]. This website provided the medical record, an educational guide, and a message system for patients. Moreover, patients could contact the health provider by email. The telemedicine diabetes disease management program Health Buddy was tested in 2 studies [33,34]. However, the modality used in the diabetes program was different from the previously mentioned monitoring programs. Patients answered personalized questions that enabled them to monitor their disease symptoms, medication adherence, and disease knowledge by pressing buttons for response. The 2 studies using the Health Buddy differed in the intensity of the feedback. One study [27] tested a medication compliance device. Data and answers to questions were recorded by the device and uploaded daily to a central server. Based on these answers health providers were able to monitor the patients, provide advice, and update the treatment regimens in the Med-eMonitor devices. One study [35] also tested a handheld device, Pocket Personal Assistant for Tracking Health (PATH), developed for patients after lung transplantation to record health data, review data trends, and report their condition changes to the transplant team. The device included decision-support programs to promote self-care behaviors.

Customized Health Programs

Interventions that provide not only tailored content but also individualized instructions for meeting certain health goals, self-management goals, or goal-setting activities are so-called customized health programs [7], used by 4 of the included studies. Of these, 2 studies [36,37] tested the effects of an Internet-based self-management program for asthma patients. This website allowed monitoring through the website, text messages, use of an Internet-based treatment plan, online education, and the possibility to communicate with the health provider. The intervention WebEase [38] consisted of three modules that were designed to assess an individual's status related to self-management practices and create a plan for change or to maintain the behavior. The modules in WebEase required the patient to answer questions related to these topics. Feedback was provided based on these responses. Patients entered data into MyLog, which is a screen for recording data about medication-taking behavior, stress, etc. In addition, the intervention included a knowledge component and a discussion board. This means that each patient was directed to another path [38]. Another study [39] tested a customized health program where patients chose which components of the website they wanted to use. The website included a home page, posttransplant skills workshops, discussion groups, "ask an expert," question-and-answer possibility, healthy-living tips, resources, and reference library. The way the patients used the website was based more on voluntary participation than in the study that used MyLog [38].

Table 4 and Table 5 show for each study which method for delivering the tailored message was used (see column 2).

Table 4. Effectiveness of short-term interventions (<6 months).

Study	Study quality	Sophistication of tailoring	Quality measurement adherence	Short-term effectiveness (<6 months)[a]
DeVito Dabbs [35]	High	Moderate	Low	++
Jan [28]	High	Moderate	Low	++
Dew [39]	Low	High	Low	–
Dilorio [38]	Low	High	Low	++
Artinian [27]	Low	Moderate	Low	+
Guendelman [34]	Low	Moderate	Low	++

[a] ++ = significant effect on medication adherence; + = moderate effect on medication adherence; – = no effect on medication adherence.

XSL•FO
RenderX

Table 5. Effectiveness of long-term interventions (>6 months).

Study	Study quality	Sophistication of tailoring	Quality measurement adherence	Long-term effectiveness (>6 months)[a]
Van der Meer [36]	High	High	Low	+
Van der Meer [37]	High	High	Low	+
Chan [30]	High	Moderate	High	+
Joseph [31]	High	Moderate	Low	++
Ross [32]	High	Moderate	Low	+
Chan [29]	Low	Moderate	Low	–
Cherry [33]	Low	Moderate	Low	+

[a] ++ = significant effect on medication adherence; + = moderate effect on medication adherence; – = no effect on medication adherence.

Role of Health Providers

Interventions also differ in the type and extent of health provider involvement. User control allows individuals to take a major role in managing their own care, whereas in expert control an expert or therapist takes a more directive role [7].

Only 4 of the interventions were based on user control and 9 interventions used contact with the health provider. A Web-based asthma management program was developed in 1 study [31]. The program used tailoring to alter behavior through individualized health messages based on the user's beliefs, attitudes, and personal barriers to change. The health provider did not interfere. Three studies were also user based with treatment algorithms to give feedback [36-38].

In contrast, in the Blue Angel for Asthma Kids, a customized educational and monitoring website site providing secure email contact between patients and their therapist, the therapist had a more directive role [28]. Like the Blue Angel for Asthma Kids, SPPARO included a messaging system that made it possible to exchange secure messages with the health provider [32]. The intervention developed in 2 studies [29,30] consisted of a case manager who reviewed the data, sent emails about the peak flow, inhaler technique, and symptoms, and forwarded them the website. Patients (the virtual group as well as the office-based group) had access to their case manager 24 hours a day, 7 days a week. One study [39] conducted a website including skills workshops, discussion group, ask an expert, question and answer, health tips, recourses, and references. In this case, the role of health providers was to provide the possibility for a patient to ask an expert.

Several interventions used a special device to exchange information between patients and health care providers. The Med-eMonitor recorded data and answers to questions that could be uploaded by the health provider. Based on these outcomes, the health provider provided advice and updated the treatment regimens [27]. The PATH [35] recorded data and provided a tailored decision-support program and email contact with the health provider. The Health Buddy device in 2 studies [33,34] is like the Med-eMonitor and PATH based on the feedback and participation of the health provider, but there is a difference in the intensity of participation of the health provider. In 1 study [33] the health provider contacted the patient only when necessary, while in the other 2 studies [34,35] the patient received feedback instantly after sending a question.

Our analysis to examine the extent to which medication adherence is determined by different tailoring levels revealed no clear relationship between the intervention's level of sophisticated tailoring and the extent to which the intervention was effective (Table 4).

Summary of Effects on Adherence

We found 5 studies with a significant effect on adherence (3 HQ studies and 2 LQ studies) [28,31,34,35,38]. The first study [28] concluded that the intervention had a positive and significant effect on use of the inhaled corticosteroid and that this effect significantly differed from the baseline. In addition, the second study [31] found positive changes in controller medication adherence. The third study [38] tested the WebEase intervention and found a significant effect on adherence. The fourth study [34] found that patients were more likely to take their asthma medication when they used the Health Buddy. In the fifth study [35], patients who received the PATH intervention were more likely to adhere to their medical regimen.

A moderate effect on adherence was reported in 6 studies (4 HQ studies and 2 LQ studies). The SPPARO proved to be feasible and improved general adherence. Adherence to medication showed a similar trend but these results did not reach significance [32]. The second study [30] concluded that, after the intervention, the use of beta-agonist decreased, which is an indication of better adherence. The third study [33] reported that medication compliance improved from 65% at pretest to 94% at posttest, but the difference was not statistically tested. The medication compliance rate in the fourth study [27] was 94% for the monitor group as measured by the monitor system. However, because there was no pretest and the data of the control group were not available, it is unknown whether the results were significant as compared with the pretest or the control group. The intervention of the last 2 studies [36,37] improved adherence in patients with partly and uncontrolled asthma at baseline. The authors concluded that the intervention was most effective in improving adherence for patients with partly or uncontrolled asthma at baseline.

No significant results on patients' adherence were found in 2 studies (1 HQ and 1 LQ) [29,39].

XSL•FO
RenderX

Assessment of Adherence Measurements

Regarding the measurement of adherence, the 13 studies we reviewed showed a large variability of methods: 12 studies used a low-quality measurement to assess adherence and 1 used a combination of these methods (ie, a high-quality measurement to assess adherence) (Table 3).

Low Quality of Adherence Measurement

In 10 studies, self-reported scales were used to obtain the adherence rate. Although 5 studies [28,31,34,36,37] used self-reports to measure adherence, they did not describe what kind of instrument they used. One study [39] used self-reported data by asking questions regarding adherence during the initial interview. Reports of therapist and patients were compared. In addition, 1 study [33] used a self-developed medication adherence survey on the Health Buddy appliance. The other 3 studies chose existing, valid, self-reported adherence scales. One study [32] used a combination of the Morisky scale and the General Adherence Scale from the Medical Outcomes Study, and 1 study [35] used the Health Habits Assessment, a self-reported scale to measure adherence. One study [38] used the self-report USCF Adherence Questionnaire and the Antiretroviral General Adherence Scale. Finally, 2 studies used measurements such as counting pills [27], and 1 study [29] used a computerized prescription refill record (after evaluation of the pilot study in which completing the diary turned out to be time consuming and inconvenient; see [30]).

High Quality of Adherence Measurement

One study used a combination of methods. This study [30] used a diary in combination with a computerized prescription refill record.

Table 3 shows the results of the assessment of the internal validity and the quality of adherence measurement.

Relation Between Quality of Adherence Measurement and Effectiveness

Our investigation of the relationship between the quality of the adherence measurement and the effectiveness of the interventions revealed no clear relationship (there was only 1 study using a high-quality method to assess adherence), although self-reported adherence measurements seemed to result more often in significant effects than did pill counts and pharmacist adherence measurements (Table 4 and Table 5). Of the 10 studies using self-reports (low-quality adherence measurement), 5 reported a significant effect of the intervention on adherence [28,31,34,35,38], 4 a moderate effect [32,33,36,37], and 1 no effect [39]. From the 2 studies in which pharmacist data or pill counting was used, 1 reported a moderate effect [27] and 1 no effect [29]. The 1 study that used a combination of methods to measure adherence [30] found a moderate effect on adherence.

Relation Between Interval of Adherence Measurement and Effectiveness

There was no clear relationship between the timing of the adherence measurements and the effectiveness of the intervention. The intervals between baseline and follow-up measurements differed between projects. Short-term adherence (ie, within 6 months) was measured in 6 studies. The first study [38] showed that WebEase improved adherence 6 weeks after baseline. Patients who used PATH were more likely to show higher adherence than the control group after an interval of 8 weeks after baseline [35]. The third study [28] found a significant effect on adherence after 12 weeks, and the fourth study [34] reported an improvement in adherence after 12 weeks. The fifth study [27] found an adherence rate of 94% in the experimental group after 12 weeks. Because of the lack of adherence data for the control group and the lack of a pretest, the effects on adherence could not be established. The sixth study [39] examined the proportion of nonadherent patients in both an intervention and a control group after 16 weeks. That study's authors found uniformly small and nonsignificant differences between the control and intervention groups. However, they found an important difference within the intervention group. Subgroup differences appeared when the intensity of using parts of the intervention was related to the effectiveness. For example, patients who used the Managing Medical Regimen Workshop more often or intensely appeared to be more adherent than those using the intervention less often or intensely [39].

Long-term adherence was measured in 7 studies—that is, adherence with an interval of 6 months or longer, mostly of 1 year or more. Two studies [30,36] reported a moderate effect in their pilot on adherence after 6 months. In 2 studies [31,32], they found a moderate [32] and significant [31] improvement in adherence after 1 year. Two studies [33,37] found a moderate effect on adherence after 1 year. This means that all of the included studies using an interval of 6 months or longer showed an effect (significant or moderate) on long-term adherence. One study [29] did not find an effect on adherence after 1 year.

Table 4 and Table 5 give an overview of the methodological quality of the studies, the level of sophistication of each intervention, the quality of measurement of adherence, and an overview of the short-term and long-term effects.

Data Synthesis

Using the principles of the best-evidence synthesis (see Textbox 1), taking into account the design, methodological quality, and outcomes of the studies, the following conclusions can be drawn. In total, 7 studies were considered HQ. We found 3 HQ studies [28,31,35] and 2 LQ studies [34,38] that had a significant effect on enhancing medication adherence and that met 6 of 11 criteria. This means that there is evidence that tailored Internet interventions are successful in improving medication adherence.

Sensitivity Analysis

The sensitivity analysis showed the same results as the best-evidence synthesis. The results remained the same when the analysis was repeated with the 6 LQ studies excluded (ie, taking only the 7 HQ studies into account). Moreover, when studies were rated to be HQ if 4 instead of 6 criteria of interval validity were met, results stayed the same.

XSL•FO
RenderX

Discussion

Principal Results

First, our objective was to gain insight into the current state of the use of Internet interventions to improve medication adherence. Results of this review indicate that this is still a new field. This is visible in the differences in interventions with respect to crucial aspects such as the level of sophisticated tailoring and the role of health care providers. Despite the differences, it is remarkable that none of the interventions used a low level of tailoring and the majority (9 of 13) provided the opportunity to contact a health provider.

Second, the studies were assessed on their effectiveness on medication adherence. There is evidence that Internet interventions can improve medication adherence. This evidence comes from 3 HQ studies and 2 LQ studies, finding significant results on medication adherence.

Third, we wanted to investigate to what degree adherence is determined by the characteristics of the intervention. All interventions discussed in this review used tailored methods and used a moderately or highly sophisticated tailored intervention. These types of health programs, especially customized health programs, are more complex, generally long-term, allowing the patients to access the programs several times [7], and are considered appropriate for difficult-to-influence behaviors. We did not find a clear relationship between how sophisticated the tailoring of the intervention was and the extent to which the intervention appeared to be effective, possibly due to the various methods that were used.

Last, we wanted to investigate whether there is a relationship between the characteristics of a study and the reported effectiveness of the interventions. We found that there was variation not only in the level of tailoring, but also in the measurement of adherence, the timing of measuring adherence, and the intensity of the intervention. The included studies used self-reporting measurements (ie, interviews, diary, self-reporting via questionnaires) or pill counts, or prescription refills, or a combination. No study used electronic monitoring, which is perceived as a high-quality method for assessing adherence [15]. Of the 13 studies we reviewed, 7 measured long-term adherence, using an interval of 6 months or longer. There is no clear evidence that the duration of the intervention is related to the effectiveness of the intervention. Nevertheless, of the 7 studies measuring long-term adherence, 1 HQ study showed positive effects and 4 HQ studies and 1 LQ study showed moderate effects on adherence. This indicates that long-term interventions are promising. However, more research in this field is needed.

There is evidence that Internet interventions can be effective in improving medication adherence. The evidence comes from 3 HQ studies. However, the results should be interpreted with caution. Self-reported scales were used in 10 studies, which is considered a low-quality adherence measurement: 5 reported a significant effect of the intervention on adherence, 4 a moderate effect, and 1 no effect. Self-measurements can contribute to overestimating of the effects of interventions [40]. This could be explained by the possibility that patients may forget that they missed a dose. Biases that appear most prominent in estimating adherence by the patient from structured questionnaires are social desirability and social approval [5,13]. In other words, studies relying on self-reporting may have a tendency to err on the optimistic side when it comes to adherence, certainly compared with more objective pill-counting studies. However, it must be noted that anonymous self-report questionnaires are found to be significantly correlated with electronic monitoring [41] and virologic response [40], considered more objective methods. Research also shows that using specific strategies, such as ensuring patients that their responses will be kept confidential [40] or stratifying patients according to their socially desirable response [42], improves the prediction of adherence by self-reports. This indicates that self-reports are not useless, but future research should examine more strategies to reinforce accurate reporting by patients [40]. On the other hand, we included RCTs and, consequently, self-reported adherence can be expected to be overestimated in both treatment arms. Thus, the intervention effect (ie, difference between intervention and control group) was not necessarily overestimated. Additionally, a distinction can be made between valid self-report measurements and measurements that are not. If self-reported measurements are used, for instance because this is a cost-effective method, using validated measurements is recommended.

Electronic monitoring or observation is considered to be more accurate. It electronically records the time and date of the actual dosing events. Because every single method has its limitations, the best approach is to use multiple assessment techniques concurrently, as a way to improve the accuracy of adherence assessment [14]. One study in our review [30] used self-reported diary and prescription refills to measure adherence and made a distinction in interpreting the results. They used a self-reported diary to assess how the patients used their inhaler, based on the idea that when patients are not using their inhaler according to the health providers' advice, they can be considered nonadherent. In addition, they used prescription refills to measure how many refills the patient was obtaining. In line with this method of measuring adherence, the optimal approach could be suggested to be a combination of self-reports and more objective measurements [43]. In addition, every single measurement needs a different interpretive approach because it has different relationships to clinical outcomes [13].

Of the studies we reviewed, 6 measured short-term adherence that varied from 6 weeks to 4 months: 4 of them were found highly effective and 1 moderately effective regarding adherence. The question is whether adherence improved in the long term, because the period was too short to measure persistence. According to an international expert forum on adherence [13], it is not easy to identify adherent and nonadherent patients beforehand. There is a large body of evidence dominated by reports identifying factors that are predictive or associated with nonadherence. Adherence could be seen as a dynamic behavior that is determined or influenced by unrelated factors that fluctuate and change over time [44]. As an adherent patient can become nonadherent over time, the importance of time (ie,

XSL•FO
RenderX

persistence) has been emphasized [5]. The quality of execution of the treatment plan can influence persistence. Factors such as perceptions of treatment outcomes, beneficial effects, and adverse effects can influence the quality of execution over time. Therefore, conducting interventions that address long-term adherence and overcome reasons why a patient is not able or not willing to adhere are recommended.

Methodological Limitations

The search method was top-down in that we relied on existing databases and search terms. This approach has the possibility of missing important articles due to miscoding of search terms. A bottom-up strategy is more time consuming but has the advantage of being more comprehensive.

Clinical Implications

Monitoring adherence to optimize effects and minimize nonadherence could be time consuming. Computers, however, are very good at collecting data concerning the monitoring of adherence. Internet interventions can be tailored, collect data, and monitor adherence. In addition, based on this systematic literature review, there is evidence that tailored Internet interventions can be an effective method to improve medication adherence. This means that Web-based interventions can be effective at increasing medication adherence among chronically ill patients. Health providers, who want to enhance patients' adherence, are encouraged to use tailored websites or reminder systems. They could use these interventions in addition to their everyday work.

Implications for Research

Because we did not find a clear relationship between the effectiveness and the degree of tailoring, we recommend that future studies should be conducted with variation in the level of sophistication of tailoring to further test which characteristics of the tailored messages have the most positive effects on adherence. Moreover, website compliance is often not completely reported. While the frequency in which the patient used the website is often reported, studies do not describe how exactly patients used the website. Therefore, it is difficult to compare the results of different interventions, because the way patients use the website can have implications for the effectiveness of the website.

Conclusion

With more than 40 million people using the Internet for a variety of purposes, health communication programs in the future are more likely to be delivered online [6]. These types of interventions especially have the potential to address difficult-to-change behaviors such as adherence. This review shows promising results on the effectiveness of tailored Internet interventions to enhance medication adherence of chronically ill patients. There is evidence that these interventions can enhance adherence. But it remains a relatively new field, and studies using more objective measurements to assess adherence are recommended.

Acknowledgments

The authors acknowledge the support of 2Comply. 2Comply had no role in the design and the collection, analysis and interpretation of the data, nor in the writing of this manuscript.

Conflicts of Interest

None declared

Authors' Contributions

JvW and LvD designed the study and acquired funding. MV conducted comprehensive literature searches through the databases. Two researchers pairs (AL/JvW and MV/LvD) performed a first selection for inclusion based on the abstracts. After having obtained full text of articles that seemed to meet the inclusion criteria, the same researchers pairs independently reviewed whether these studies fulfilled all the inclusion criteria. The methodological quality of included RCT's and CCT's were independently reviewed by AL and JvW. In addition, AL and LvD independently assessed the quality of the methods for measuring adherence to a medical regime. AL conducted the 'Best Evidence Synthesis' and drafted the manuscript. MV, LvD, ES and JvW have been involved in critically revising the manuscript. All authors have read and approved the final manuscript.

Multimedia Appendix 1

Detailed description of the excluded studies.

[PDF file (Adobe PDF File), 127 KB - jmir_v13i4e103_app1.pdf]

Multimedia Appendix 2

Detailed description of the characteristics of the included studies.

[PDF file (Adobe PDF File), 155 KB - jmir_v13i4e103_app2.pdf]

References

XSL•FO
RenderX

1. Sabaté E. Adherence to Long-Term Therapies: Evidence for Action. Geneva: World Health Organization; 2003.
2. 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]
3. Kane SV, Robinson A. Review article: understanding adherence to medication in ulcerative colitis - innovative thinking and evolving concepts. Aliment Pharmacol Ther 2010 Nov;32(9):1051-1058. [doi: 10.1111/j.1365-2036.2010.04445.x] [Medline: 20815833]
4. Wroe AL. Intentional and unintentional nonadherence: a study of decision making. J Behav Med 2002 Aug;25(4):355-372. [Medline: 12136497]
5. Urquhart J, Vrijens B. New findings about patient adherence to prescribed drug dosing regimes: an introduction to pharmionics. Eur J Hosp Pharm Sci 2005;11(5):103-106.
6. Kreuter MW, Farrell D, Olevitch L, Brenna L. Tailoring Health Messages: Customizing Communication With Computer Technology. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
7. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
8. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
9. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003;27 Suppl 3:S227-S232. [Medline: 14672383]
10. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
11. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions, version 5.1.0.: The Cochrane Collaboration; 2011 Mar. URL: http://www.cochrane-handbook.org/ [accessed 2011-10-07] [WebCite Cache ID 63dFDmjXv]
12. van Tulder M, Furlan A, Bombardier C, Bouter L, Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine (Phila Pa 1976) 2003 Jun 15;28(12):1290-1299. [doi: 10.1097/01.BRS.0000065484.95996.AF] [Medline: 12811274]
13. Sluijs E, van Dulmen S. van DijkL, Ridder D, Heerink R, Bensing J. Patient Adherence to Medical Treatment: A Meta Review. Utrecht: NIVEL; 2006.
14. Wetzels GE, Nelemans PJ, Schouten JS, van Wijk BL, Prins MH. All that glisters is not gold: a comparison of electronic monitoring versus filled prescriptions--an observational study. BMC Health Serv Res 2006;6:8 [FREE Full text] [doi: 10.1186/1472-6963-6-8] [Medline: 16472388]
15. Farmer KC. Methods for measuring and monitoring medication regimen adherence in clinical trials and clinical practice. Clin Ther 1999 Jun;21(6):1074-90; discussion 1073. [doi: 10.1016/S0149-2918(99)80026-5] [Medline: 10440628]
16. Velligan DI, Wang M, Diamond P, Glahn DC, Castillo D, Bendle S, et al. Relationships among subjective and objective measures of adherence to oral antipsychotic medications. Psychiatr Serv 2007 Sep;58(9):1187-1192 [FREE Full text] [doi: 10.1176/appi.ps.58.9.1187] [Medline: 17766564]
17. Pullar T, Kumar S, Tindall H, Feely M. Time to stop counting the tablets? Clin Pharmacol Ther 1989 Aug;46(2):163-168. [Medline: 2758726]
18. Rudd P, Byyny RL, Zachary V, LoVerde ME, Titus C, Mitchell WD, et al. The natural history of medication compliance in a drug trial: limitations of pill counts. Clin Pharmacol Ther 1989 Aug;46(2):169-176. [Medline: 2667837]
19. Shi L, Liu J, Fonseca V, Walker P, Kalsekar A, Pawaskar M. Correlation between adherence rates measured by MEMS and self-reported questionnaires: a meta-analysis. Health Qual Life Outcomes 2010;8:99 [FREE Full text] [doi: 10.1186/1477-7525-8-99] [Medline: 20836888]
20. Shi L, Liu J, Koleva Y, Fonseca V, Kalsekar A, Pawaskar M. Concordance of adherence measurement using self-reported adherence questionnaires and medication monitoring devices. Pharmacoeconomics 2010;28(12):1097-1107. [doi: 10.2165/11537400-000000000-00000] [Medline: 21080735]
21. Grymonpre RE, Didur CD, Montgomery PR, Sitar DS. Pill count, self-report, and pharmacy claims data to measure medication adherence in the elderly. Ann Pharmacother 1998;32(7-8):749-754. [Medline: 9681089]
22. DiMatteo MR, Giordani PJ, Lepper HS, Croghan TW. Patient adherence and medical treatment outcomes: a meta-analysis. Med Care 2002 Sep;40(9):794-811. [doi: 10.1097/01.MLR.0000024612.61915.2D] [Medline: 12218770]
23. Sajatovic M, Velligan DI, Weiden PJ, Valenstein MA, Ogedegbe G. Measurement of psychiatric treatment adherence. J Psychosom Res 2010 Dec;69(6):591-599. [doi: 10.1016/j.jpsychores.2009.05.007] [Medline: 21109048]
24. McMahon JH, Jordan MR, Kelley K, Bertagnolio S, Hong SY, Wanke CA, et al. Pharmacy adherence measures to assess adherence to antiretroviral therapy: review of the literature and implications for treatment monitoring. Clin Infect Dis 2011 Feb 15;52(4):493-506. [doi: 10.1093/cid/ciq167] [Medline: 21245156]
25. Steultjens EMJ, Dekker J, Bouter LM. van de Nes JCM, Cardol M, van den Ende CHM. Occupational Therapy for Multiple Sclerosis (Protocol for a Cochrane Review). Oxford: Update Software; 2002.

XSL•FO
RenderX

26. Verkaik R, van Weert JC, Francke AL. The effects of psychosocial methods on depressed, aggressive and apathetic behaviors of people with dementia: a systematic review. Int J Geriatr Psychiatry 2005 Apr;20(4):301-314. [doi: 10.1002/gps.1279] [Medline: 15799081]
27. Artinian NT, Harden JK, Kronenberg MW, Vander Wal JS, Daher E, Stephens Q, et al. Pilot study of a Web-based compliance monitoring device for patients with congestive heart failure. Heart Lung 2003 Aug;32(4):226-233. [Medline: 12891162]
28. Jan RL, Wang JY, Huang MC, Tseng SM, Su HJ, Liu LF. An internet-based interactive telemonitoring system for improving childhood asthma outcomes in Taiwan. Telemed J E Health 2007 Jun;13(3):257-268. [doi: 10.1089/tmj.2006.0053] [Medline: 17603828]
29. Chan DS, Callahan CW, Hatch-Pigott VB, Lawless A, Proffitt HL, Manning NE, et al. Internet-based home monitoring and education of children with asthma is comparable to ideal office-based care: results of a 1-year asthma in-home monitoring trial. Pediatrics 2007 Mar;119(3):569-578 [FREE Full text] [doi: 10.1542/peds.2006-1884] [Medline: 17332210]
30. Chan DS, Callahan CW, Sheets SJ, Moreno CN, Malone FJ. An Internet-based store-and-forward video home telehealth system for improving asthma outcomes in children. Am J Health Syst Pharm 2003 Oct 1;60(19):1976-1981. [Medline: 14531243]
31. Joseph CL, Peterson E, Havstad S, Johnson CC, Hoerauf S, Stringer S, Asthma in Adolescents Research Team. A web-based, tailored asthma management program for urban African-American high school students. Am J Respir Crit Care Med 2007 May 1;175(9):888-895 [FREE Full text] [doi: 10.1164/rccm.200608-1244OC] [Medline: 17290041]
32. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin CT. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]
33. Cherry JC, Moffatt TP, Rodriguez C, Dryden K. Diabetes disease management program for an indigent population empowered by telemedicine technology. Diabetes Technol Ther 2002;4(6):783-791. [Medline: 12685802]
34. Guendelman S, Meade K, Benson M, Chen YQ, Samuels S. Improving asthma outcomes and self-management behaviors of inner-city children: a randomized trial of the Health Buddy interactive device and an asthma diary. Arch Pediatr Adolesc Med 2002 Feb;156(2):114-120 [FREE Full text] [Medline: 11814370]
35. DeVito Dabbs A, Dew MA, Myers B, Begey A, Hawkins R, Ren D, et al. Evaluation of a hand-held, computer-based intervention to promote early self-care behaviors after lung transplant. Clin Transplant 2009;23(4):537-545. [doi: 10.1111/j.1399-0012.2009.00992.x] [Medline: 19473201]
36. van der Meer V, Bakker MJ, van den Hout WB, Rabe KF, Sterk PJ, Kievit J, SMASHING (Self-Management in Asthma Supported by Hospitals‚ ICT‚ Nurses and General Practitioners) Study Group. Internet-based self-management plus education compared with usual care in asthma: a randomized trial. Ann Intern Med 2009 Jul 21;151(2):110-120. [Medline: 19620163]
37. van der Meer V, van Stel HF, Bakker MJ, Roldaan AC, Assendelft WJ, Sterk PJ, SMASHING (Self-Management of Asthma Supported by Hospitals‚ ICT‚ Nurses and General practitioners) Study Group. Weekly self-monitoring and treatment adjustment benefit patients with partly controlled and uncontrolled asthma: an analysis of the SMASHING study. Respir Res 2010;11:74 [FREE Full text] [doi: 10.1186/1465-9921-11-74] [Medline: 20537124]
38. Dilorio C, Escoffery C, McCarty F, Yeager KA, Henry TR, Koganti A, et al. Evaluation of WebEase: an epilepsy self-management Web site. Health Educ Res 2009 Apr;24(2):185-197 [FREE Full text] [doi: 10.1093/her/cyn012] [Medline: 18682382]
39. Dew MA, Goycoolea JM, Harris RC, Lee A, Zomak R, Dunbar-Jacob J, et al. An internet-based intervention to improve psychosocial outcomes in heart transplant recipients and family caregivers: development and evaluation. J Heart Lung Transplant 2004 Jun;23(6):745-758. [Medline: 15366436]
40. Nieuwkerk PT, Oort FJ. Self-reported adherence to antiretroviral therapy for HIV-1 infection and virologic treatment response: a meta-analysis. J Acquir Immune Defic Syndr 2005 Apr 1;38(4):445-448. [Medline: 15764962]
41. Hugen PW, Langebeek N, Burger DM, Zomer B, van Leusen R, Schuurman R, et al. Assessment of adherence to HIV protease inhibitors: comparison and combination of various methods, including MEMS (electronic monitoring), patient and nurse report, and therapeutic drug monitoring. J Acquir Immune Defic Syndr 2002 Jul 1;30(3):324-334. [Medline: 12131570]
42. Nieuwkerk PT, de Boer-van der Kolk IM, Prins JM, Locadia M, Sprangers MA. Self-reported adherence is more predictive of virological treatment response among patients with a lower tendency towards socially desirable responding. Antivir Ther 2010;15(6):913-916. [doi: 10.3851/IMP1644] [Medline: 20834104]
43. Drotar D, Richard KA, Burgess E, Levi R, Nobile C, Seti A, et al. Adherence to treatment in childhood chronic illness: issues and recommendations to enhance pediatric research and training. In: Drotar D, editor. Promoting Adherence to Treatment in Childhood Chronic Illness: Concepts, Methods and Interventions. Mahwah, NJ: Erlbaum; 2002.
44. Reynolds NR. Adherence to antiretroviral therapies: state of the science. Curr HIV Res 2004 Jul;2(3):207-214. [Medline: 15279584]

XSL•FO
RenderX

Abbreviations

CCT: clinical controlled trial
HQ: high quality (study)
LQ: low quality (study)
PATH: Pocket Personal Assistant for Tracking Health
RCT: randomized controlled trial
SPPARO: System Providing Access to Records Online

Edited by G Eysenbach; submitted 24.01.11; peer-reviewed by P Nieuwkerk, C Stepnowsky; comments to author 25.04.11; revised version received 19.07.11; accepted 19.07.11; published 05.12.11.

Please cite as:
Linn AJ, Vervloet M, van Dijk L, Smit EG, Van Weert JCM
Effects of eHealth Interventions on Medication Adherence: A Systematic Review of the Literature
J Med Internet Res 2011;13(4):e103
URL: http://www.jmir.org/2011/4/e103/
doi:10.2196/jmir.1738
PMID:22138112

©Annemiek J Linn, Marcia Vervloet, Liset van Dijk, Edith G Smit, Julia CM Van Weert. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of a Web-Based Intervention to Promote Hand Hygiene: Exploratory Randomized Controlled Trial

Lucy Yardley[1], PhD; Sascha Miller[1], MSc; Wolff Schlotz[2], PhD; Paul Little[3], MD, PhD, FRCGP

[1]Academic Unit of Psychology, Faculty of Human and Social Sciences, University of Southampton, Southampton, United Kingdom

[2]Institute of Experimental Psychology, University of Regensburg, Regensburg, Germany

[3]Primary Care and Population Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Lucy Yardley, PhD
Academic Unit of Psychology
Faculty of Human and Social Sciences
University of Southampton
Highfield
Southampton, SO17 1BJ
United Kingdom
Phone: 44 2380 59 4581
Fax: 44 2380 59 4597
Email: L.Yardley@soton.ac.uk

Abstract

Background: Hand-washing is regarded as a potentially important behavior for preventing transmission of respiratory infection, particularly during a pandemic.

Objective: The objective of our study was to evaluate whether a Web-based intervention can encourage more frequent hand-washing in the home, and to examine potential mediators and moderators of outcomes, as a necessary first step before testing effects of the intervention on infection rates in the PRIMIT trial (PRimary care trial of a website based Infection control intervention to Modify Influenza-like illness and respiratory infection Transmission).

Methods: In a parallel-group pragmatic exploratory trial design, 517 nonblinded adults recruited through primary care were automatically randomly assigned to a fully automated intervention comprising 4 sessions of tailored motivational messages and self-regulation support (n = 324) or to a no-intervention control group (n = 179; ratio 2:1). Hand-washing frequency and theory of planned behavior cognitions relating to hand-washing were assessed by online questionnaires at baseline (in only half of the control participants, to permit evaluation of effects of baseline assessment on effect sizes), at 4 weeks (postintervention; all participants), and at 12 weeks.

Results: Hand-washing rates in the intervention group were higher at 4 weeks than in the control group (mean 4.40, n = 285 and mean 4.04, n = 157, respectively; $P < .001$, Cohen $d = 0.42$) and remained higher at 12 weeks (mean 4.45, n = 282 and mean 4.12, n = 154, respectively; $P < .001$, Cohen $d = 0.34$). Hand-washing intentions and positive attitudes toward hand-washing increased more from baseline to 4 weeks in the intervention group than in the control group. Mediation analyses revealed positive indirect effects of the intervention on change in hand-washing via intentions (coefficient = .15, 95% confidence interval [CI], .08–.26) and attitudes (coefficient = 0.16, 95% CI, .09–.26). Moderator analyses confirmed that the intervention was similarly effective for men and women, those of higher and lower socioeconomic status, and those with higher and lower levels of perceived risk.

Conclusions: This study provides promising evidence that Web-based interventions could potentially provide an effective method of promoting hand hygiene in the home. Data were collected during the 2010 influenza pandemic, when participants in both groups had already been exposed to extensive publicity about the need for hand hygiene, suggesting that our intervention could add to existing public health campaigns. However, further research is required to determine the effects of the intervention on actual infection rates.

Trial: International Standard Randomized Controlled Trial Number (ISRCTN): 75058295; http://www.controlled-trials.com/ISRCTN75058295 (Archived by WebCite at http://www.webcitation.org/62KSbkNmm)

XSL•FO
RenderX

(J Med Internet Res 2011;13(4):e107) doi:10.2196/jmir.1963

KEYWORDS

Health promotion; human influenza; hand-washing; randomized controlled trial

Introduction

Respiratory infections, such as influenza and even the common cold, continue to present a major health problem in the 21st century. Influenza pandemics have the potential to cause substantial morbidity and mortality as well as widespread social and economic disruption [1]. While the 2009 H1N1 pandemic proved relatively mild for most people, a much more severe influenza pandemic (eg, H5N1) is still anticipated, which could result in many millions of deaths worldwide [2]. In nonpandemic years, colds and influenza still pose a considerable burden for individuals, health services, and society through their impact on quality of life, the ability to work, vulnerability to more serious illness, and need for medical care [3,4].

The relative importance of different routes of infection by influenza has not yet been established, but the current consensus is that transmission from hand to face could play a significant role [5]. Adoption of simple preventive hygiene measures, especially frequent hand-washing, could prove a cost-effective means of reducing transmission of respiratory infections [6-9], and these measures were therefore recommended during the H1N1 pandemic by the World Health Organization and promoted in national campaigns worldwide. Slowing the spread of infection could help to prevent health and other services from becoming overwhelmed and allow time for the development and distribution of vaccines [10]. However, surveys carried out in the context of both severe acute respiratory syndrome (SARS) and influenza pandemics have found that less than half of those surveyed reported adhering to recommended rates of hand-washing (at least 10 times a day), in both community and higher-risk samples [11-14]. Adherence to hygiene recommendations is probably lower than these surveys suggest, since self-reported hand-washing rates typically overestimate actual hand-washing behavior [15].

There is clearly a need to develop interventions to promote hygienic behavior and test their effectiveness. Interventions are required that could be made available to the general public rapidly and at low cost, since most of the population is likely to be at risk from pandemic influenza [1]. The Internet seems an ideal medium for such an intervention; in a survey carried out in the United States, most respondents stated that the Internet would be the first source of information that they would consult in the event of a pandemic [16]. However, we are aware of only one small study of a Web-based intervention to reduce transmission of influenza [17], which found positive trends in behavior but no significant effect on hand hygiene.

When developing public health interventions, whether online or offline, it is important not only to demonstrate effectiveness but also to establish what sectors of the population can be reached by each type of intervention employed, and in particular to ensure that interventions reach those most in need of them [18,19]. While the Internet may be the best medium for reaching much of the population, it may be less effective for some sectors, such as older people and socially deprived groups [20]. Previous pandemics (including the recent H1N1 pandemic) have had a more severe impact on these sectors of the population, which are typically more vulnerable to health problems [21-24]. It is therefore vital to consider whether a Web-based hygiene intervention could be used to reach older and socially deprived people, or whether alternative interventions may be required. In addition, hand-washing rates are known to be lower in men and those less concerned about risk of infection [12,14,15,25,26], and so it is necessary to evaluate whether the intervention is effective in men and those with low perceived risk.

Developing and Testing the Intervention

Our Web-based intervention to promote hygienic behavior was developed following best practice for theory- and evidence-based intervention development [27-30]. The most appropriate target behaviors and the key attitudes and beliefs associated with these behaviors were identified by literature review and a series of qualitative and quantitative pilot studies [31,32]. The theory of planned behavior [33] was used as the principal theoretical framework, as it is flexible enough to be applied in a wide variety of contexts, it can be combined with other models and predictors, and there is evidence that components of the model that are amenable to change by intervention are key predictors of health-related behavior [34-36]. The theory of planned behavior proposes that any behavior is determined principally by the intention to perform that behavior. Intention is in turn determined by (1) attitude (a global evaluation of whether performing the behavior will have positive or negative outcomes), (2) subjective norm (the perception that relevant others would approve or disapprove of the individual carrying out the behavior), and (3) perceived behavioral control (the extent to which the individual feels it is easy or difficult to carry out the behavior). We therefore applied the model by constructing messages that would promote positive attitudes, subjective norms, and perceived behavioral control by encouraging participants to perceive hand-washing as effective, socially desirable, and easy to do. These were supplemented by theory-based techniques addressing perceived risk of pandemic flu [37], promoting appropriate illness perceptions [38], and supporting implementation of intended behavior [39,40]. In total, the intervention incorporated 18 of the 26 theory-based behavioral change techniques listed in a recently published taxonomy [41]. Our intervention was developed with input from all sectors of the community and was designed to be accessible and appropriate for men and women of all ages, of high and low socioeconomic status, and with a high and low perceived risk of infection [31].

The present study was designed to test the effects of our Web-based intervention on hand hygiene, as an essential precursor to a pragmatic trial of the effects on infection transmission. We hypothesized that hand-washing rates, and

XSL•FO
RenderX

intentions to wash hands more frequently in the future, would be higher in those given access to the intervention than in those who were not given access to it. We tested this prediction at 4 weeks (immediately after completing the intervention) and at 12 weeks, to check whether any increase in hand-washing was maintained. We also tested the prediction that the theory of planned behavior cognitions targeted by the intervention (ie, intentions, attitudes, subjective norm, and perceived behavioral control) would increase more from baseline in the intervention than in the control group, and that changes in cognitions would mediate changes in behavior. To examine potential variations in response to the intervention in different sectors of the population, we then analyzed the effects on hand-washing of age, gender, socioeconomic status, and perceived risk of infection. We hypothesized that there would be no moderator effects on intervention outcome, despite any baseline differences in hand-washing rates that might be found.

Methods

Design

Ethics approval was obtained from the National Research Ethics Service. In a parallel-group design, when participants initially logged on to the website, two-thirds were automatically randomly assigned by the intervention software to receive the intervention and one-third to the control condition, which received no intervention. No blinding of participants was possible, nor would it have been appropriate to our pragmatic design [42].

In pragmatic trials it is considered good practice to avoid intervening in the control group in any way that might change outcomes and therefore affect the comparison of effect sizes in the intervention and control groups [42]. If measurement of attitudes and behavior might affect outcomes [43], it is necessary to omit measurement until the intervention has been delivered. The rationale is that effects of measurement on behavior are likely to be greater in the control than in the intervention group, and will therefore lead to an underestimation of the intervention effect that would be observed if the intervention were implemented. For example, asking participants to answer questions that require them to reflect on their hand-washing behavior might influence behavior in a control group with no other intervention, but may not have any additive effect on the behavior of an intervention group that is exposed to extensive materials encouraging such reflection. However, it is considered good practice in behavioral research to control for measurement effects and to examine mediators of intervention effectiveness by comparing change in attitudes and behavior from baseline in the intervention and control groups. Since this behavioral study was designed as the precursor to a pragmatic trial, we felt it was important to satisfy both these requirements. We therefore randomly assigned our control participants to two subgroups: one received all the same measures as the intervention group, while the other completed measures only at 4 weeks and 12 weeks. This solution enabled us to estimate intervention effects in the absence of any contamination of control group behavior, but also allowed us to check that intervention effects could not be attributed to mere measurement.

Intervention

The intervention consisted of four weekly Web-based sessions, each containing new content in order to encourage repeat visits [44,45]. See Figure 1 and Multimedia Appendix 1 for illustrative screen shots, Multimedia Appendix 2 for more details of the intervention development and content, and http://www.lifeguideonline.org/player/play/primitdemo for demonstration pages from the first session (archived by WebCite at http://www.webcitation.org/634AW68U7). Session 1 (10 core pages) provided all the essential components of the intervention, including information about the medical team giving the advice (to enhance credibility); the need to prevent seasonal and pandemic flu; the link between hand-washing and virus transmission; expert recommendations for hand-washing frequency and technique; and instructions for picking up a free supply of hand gel from their local practice. Participants completed a hand-washing plan to promote intention formation with situational cueing. Tailored feedback was provided to help users improve their plan where necessary. Users were encouraged to print, sign, and post up the plan and involve other household members.

The three remaining sessions reinforced positive attitudes and norms and addressed common negative beliefs identified during piloting. Tailored feedback was given based on 3 items assessing current hand-washing frequency, agreement that hand-washing would prevent virus transmission, and perceived difficulty of carrying out the behavior. On logging on to the second session, half of the participants were randomly assigned to also receive advice (1 page per session) on how to reduce infection risk by boosting the immune system (eg, through a healthy lifestyle or taking echinacea). The purpose of this comparison was to check that risk-compensation mechanisms [46] did not lead to a reduction in hand-washing rates because advice on other methods of reducing infection had been given. Comparison of the intervention groups with and without these additional pages revealed absolutely no differences in outcomes, and so both intervention subgroups were pooled for analysis.

XSL•FO
RenderX

Figure 1. Homepage of the Internet Doctor website.



Participants and Procedure

Participants were recruited by mailed invitations to take part in a study of methods of reducing the spread of infection from colds and seasonal and pandemic flu. These were sent to 8150 people aged over 18 years randomly sampled from the lists of nine general practices in Southern England from August to October 2010 (4 months after the onset of the H1N1 pandemic), including practices in areas of high and low socioeconomic deprivation. The invitation letter (Multimedia Appendix 3) sought participation from people with home Internet access and living with at least one other household member.

After returning their signed consent forms and email address, participants were emailed a unique username and URL for logging on to the website. Participants who were allocated baseline assessments completed them online on their initial login. Following first login, participants in the intervention groups were emailed after 4 days to log in to session 2, and invitations to sessions 3 and 4 followed at 1-week intervals after login to the previous session (see Table 1 for an overview of study procedures and Multimedia Appendix 4 for the protocol). To prompt usage, two follow-up emails were sent to participants who did not log in to any session [44,45].

Table 1. Overview of study procedures

Time point	Intervention group	Control group with baseline measurement	Control group without baseline measurement
Recruitment	Informed consent; collection of personal details; initial login; randomization	Informed consent; collection of personal details; initial login; randomization	Informed consent; collection of personal details; initial login; randomization
Baseline	Assessment of hand-washing rates, theory of planned behavior cognitions, perceived risk	Assessment of hand-washing rates, theory of planned behavior cognitions, perceived risk	No assessment
Weeks 0–3	Weekly email invitations to log on to Web-based session promoting hand-washing	No intervention	No intervention
Week 4	Assessment of hand-washing rates, theory of planned behavior cognitions	Assessment of hand-washing rates, theory of planned behavior cognitions	Assessment of hand-washing rates, theory of planned behavior cognitions
Week 12	Assessment of hand-washing rates, theory of planned behavior cognitions	Assessment of hand-washing rates, theory of planned behavior cognitions	Assessment of hand-washing rates, theory of planned behavior cognitions

All participants were sent invitations to complete the assessment measures online at 4 weeks and 12 weeks after initial login (regardless of progress through the sessions). Two follow-up emails were sent for each assessment. To maximize follow-up, phone calls were made to nonresponders to the 4- and 12-week assessments to elicit responses to the primary outcome measure (hand-washing frequency).

Measures

Hand-washing frequency (explicitly defined as using soap and water or antibacterial gel) was assessed by a single item ranging from 1 (0–2 times a day) to 5 (10 or more times a day). Intentions were measured by a 3-item scale asking the respondent to indicate on a 7-point scale (from 1 = disagree strongly to 7 = agree strongly) that they intended to wash their

XSL•FO
RenderX

hands "at least 10 times a day," "more often," and "as often as possible" (alpha = .91). Self-reported frequency of hand-gel use was also assessed by a single item ranging from 1 (0–2 times a week) to 5 (10 or more times a week).

All measures of theory of planned behavior cognitions and perceived risk were also scored from 1 to 7; items were recoded for analysis where necessary so that higher scores indicate greater agreement, and summed subscale scores were divided by the number of items to allow direct comparison. All items assessing theory of planned behavior cognitions explicitly elicited views of hand-washing with soap or antibacterial gel at least 10 times a day (the key target behavior for the intervention). Attitudes were measured by 6 bipolar semantic differential scales: 3 items formed a direct measure of instrumental attitude (asking whether the target behavior was seen as useless/useful, unnecessary/necessary, or bad/good), and 3 measured affective attitude (asking whether the target behavior would make the respondent feel worried/confident, proud/embarrassed, or sensible/foolish). However, factor analysis indicated that these items clearly loaded on a single scale (alpha = .92): 2 items (alpha = .90) assessed subjective norms by measuring agreement (7 = agree strongly) that "people whose opinions matter to me" and "people I live with" would approve of the target behavior. Perceived behavioral control for carrying out the target behavior was assessed by 2 items (alpha = .95) measuring the self-efficacy ("I am confident that I could") and perceived control ("it will be possible for me") dimensions. Respondents indicated agreement with these statements (7 = agree strongly), which were preceded by "If I wanted to," to hold motivation constant [47,48].

Perceived risk of infection was assessed by agreement (7 = agree strongly) with 2 items (alpha = .90) assessing perceived likelihood of catching pandemic flu if no preventive action was taken [49]. This dimension of risk was assessed because pilot work indicated it was a better predictor of hand-washing intentions than was perceived worry about infection or perceived severity of infection [31,50].

Participants reported their gender, age, and postcode. The GeoConvert program [51] was used to estimate socioeconomic status from postcode, based on the Indices of Deprivation 2007 Lower Super Output Area Score (England), the official UK government measure of the relative socioeconomic deprivation associated with each postcode area, based on a weighted combination of 37 different indicators (a lower ranking denotes less deprivation). Website usage was analyzed by number of sessions accessed [52]. Practice staff kept a record of which participants collected their free sample of hand gel.

Statistical Analysis

The effectiveness of the intervention was tested first by a direct comparison (by independent *t* test, using Cohen *d* to assess effect size) of the primary and secondary outcome measures, hand-washing frequency and intentions, in the control and intervention groups at 4 weeks and 12 weeks, based on all participants who provided data at each time point. To examine possible measurement effects on outcomes, we repeated the between-group analyses at 4 weeks, comparing the intervention group with the control groups with and without baseline measurement. This analysis was not repeated for the 12-week follow-up since by that time point both control groups had been exposed to the measures. We powered the study to have 80% power to detect a small to medium effect size ($d = 0.35$) in the key comparison between the control and intervention groups with alpha = 0.05; this required a minimum sample size of 97 in the control group and 195 in the intervention group. We chose this effect size since effect sizes of Web-based interventions are typically quite small (though nevertheless potentially useful at a population level), but very small effects were not worth detecting, as they would not be clinically useful.

We further examined intervention effects by mixed-effects regression models for longitudinal data comparing change in intentions from baseline to 4 weeks in the control and intervention groups. Mixed-effects regression models were also employed to compare change in the theory of planned behavior cognitions between baseline and 4 weeks. Mixed-effects regression models use all available data within subjects, so that there is no need to replace missing values.

To examine whether intervention effects on behavior were a consequence of changes in cognitions, we used mediation analysis to test indirect effects of intervention on change in hand-washing behavior via changes in those cognitions that were targeted to be modified by the intervention. We estimated confidence limits of the total indirect effect by bias-corrected bootstrap confidence intervals (CI) with 1000 draws [53]. We used Mplus (version 6.11; Muthén & Muthén, Los Angeles, CA, USA) to calculate mediation models.

We then employed correlations to examine the relationship of gender, age, and socioeconomic status to hand-washing frequency and intentions at baseline. Multivariate analyses of variance (MANOVAs) were used to examine the interaction between intervention group and moderator effects on hand-washing frequency and intentions (combined) at the 4-week follow-up. Longitudinal subgroup analyses of moderator effects could not be carried out due to the resulting small control group cell sizes (since only 1 in 6 participants were randomly assigned to the control group and to complete baseline assessments).

Since many of the variables were not normally distributed, we confirmed all analyses by equivalent nonparametric tests, which gave virtually identical results. Finally, we examined the increase in hand-washing rates and intentions in those whose level of hand-washing at baseline was less than the recommended target (ie, those scoring less than 5), as this subgroup can be considered the target population for the intervention.

Results

Participant Characteristics and Study Participation

A total of 487 people completed the primary or secondary outcome measures at either baseline or follow-up, and so were included in either the cross-sectional or longitudinal analyses. Figure 2 shows the flow of participants for the primary outcome measure. Initial uptake was low (517/8150, 6.3% of those invited underwent random allocation) and few people explained their

XSL•FO
RenderX

reasons for nonparticipation. However, follow-up rates were good, with 157/179 (87.7%) control and 285/324 (88.0%) intervention group participants responding to the primary outcome measure at 4 weeks. Receipt of the intervention once allocated was also relatively good. Of the 324 of these participants who were randomly assigned to the intervention, 251 (77.5%) progressed to the second session, 219 (67.6%) completed three sessions, and 188 (58.0%) completed all four sessions.

The free hand gel was collected by 170/324 (52.5%) eligible participants. Those who collected hand gel were substantially more likely to report using hand gel at 4 weeks (t_{272} = 3.19, P = .002, d = 0.39), but as the mean frequency of hand gel use was only around 6 times a week this did not result in significantly higher rates of daily hand-washing (t_{283} = 1.36, P = .18, d = 0.16). In the intervention group, hand-washing at 4 weeks was associated with total time spent using the intervention (r = .23, P = .002) and number of sessions accessed (r = .21, P < .001).

Table 2 shows baseline characteristics in the intervention and control groups. There were no significant group differences at baseline (P > .10 for all comparisons). Nearly two-thirds of the sample were women, and the age range was 22 to 82 years. Among those for whom baseline hand-washing rates were assessed, 46.4% (189/407) of participants reported already achieving the recommended target of hand-washing at least 10 times a day.

Participants were excluded from these analyses only if they did not complete the primary outcome measure (hand-washing) at follow-up. Note that the sample analyzed longitudinally differs (see Table 3), as it includes those with missing data at follow-up (using imputation methods; see Statistical Analysis section) but not those allocated to the control group without baseline assessment. [b] Percentage of those randomly assigned to the group that were analyzed.

Table 2. Participant characteristics at baseline in the intervention and control groups[a]

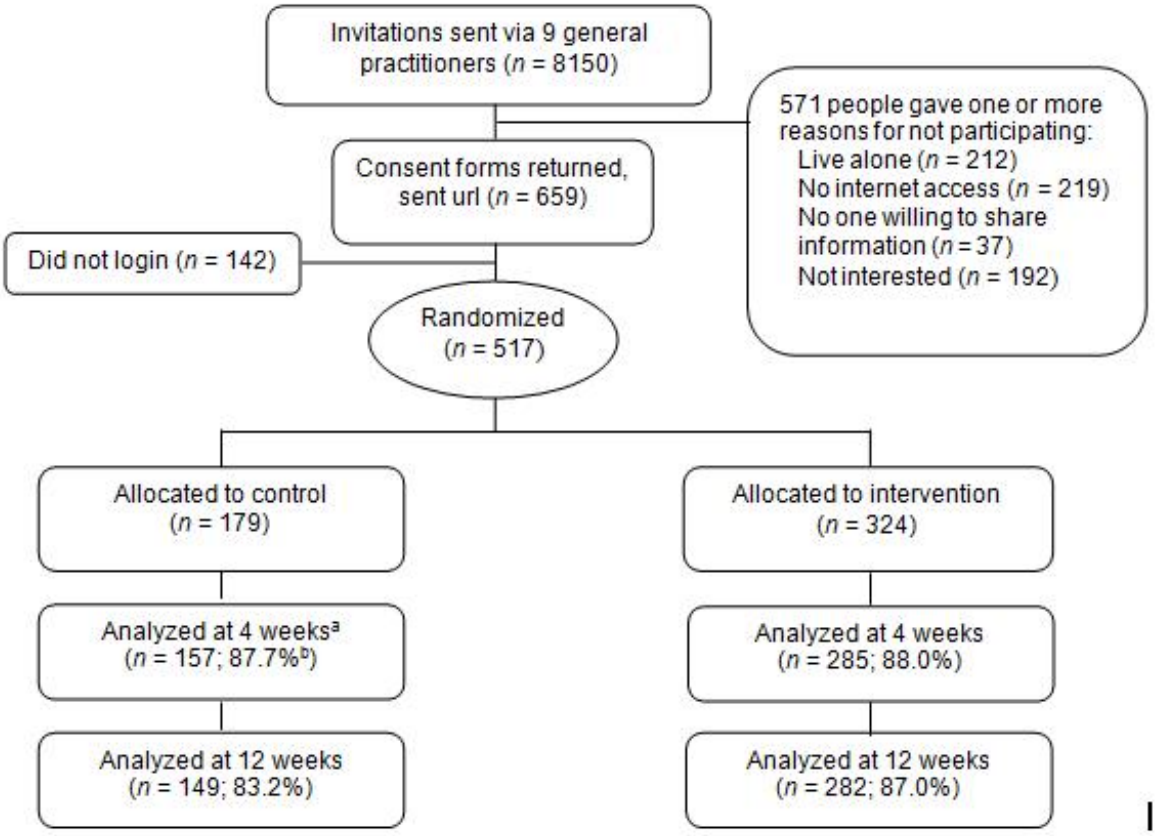
Characteristic	Intervention (n = 336)	Control (n = 181)	Total (n = 517)
Number of women (men)	213 (123)	117 (64)	330 (187)
Age (years)	49.17 (11.02)	50.94 (12.05)	49.76 (11.40)
Socioeconomic deprivation score	9.04 (6.13)	9.39 (6.88)	9.17 (6.41)
Perceived risk	5.05 (1.62)	4.77 (1.64)	4.99 (1.63)
Hand-washing frequency	4.08 (1.05)	4.01 (1.13)	4.06 (1.07)

[a]Figures are mean (SD) except where stated.

While the range of socioeconomic status observed was quite broad (1.10 to 45.10), the sample was highly skewed toward higher status, with a median of 7.87 and an interquartile range of only 5.24–11.02. Consequently, for analyses of the effects of socioeconomic status we compared those with a score less than 12 versus those with scores ranging from 12 to 45. The median risk score in the sample was 5, and so for analyses of the effects of risk we compared those with a score of 5 or more (indicating some agreement that they were likely to catch pandemic flu) with those with scores below 5.

XSL•FO
RenderX

Figure 2. Participant flow chart for primary outcome measure (hand-washing rate).



Intervention Effects

As predicted, hand-washing rates were higher postintervention in the intervention than in the control group; the key comparison of hand-washing rates and intentions in the control and intervention groups was highly significant ($P < .001$) for both measures, at 4 weeks and at 12 weeks (see Table 3). Hand-washing rates were also significantly higher in the intervention group than in the control group with baseline measurement ($t_{360} = 2.28$, $P = .02$, $d = 0.31$; mean group difference = 0.30, 95% CI, 0.04–0.55), although the effect size of the intervention was somewhat greater than in those without baseline measurement ($t_{363} = 3.41$, $P = .001$, $d = 0.45$; mean group difference = 0.43, 95% CI, 0.67–0.18).

Table 3. Between-group comparisons of hand-washing frequency and hand-washing intentions at baseline, 4 weeks, and 12 weeks[a]

Variable	Control group		Intervention group		Difference between groups, mean (95% CI[b])	Effect size, Cohen *d*
	n	Mean (SD)	n	Mean (SD)		
Hand-washing at baseline	91	4.01 (1.13)	316	4.08 (1.05)	0.06 (–0.20 to 0.33)	0.06
Hand-washing at 4 weeks	157	4.04 (0.86)	285	4.40 (0.86)	0.36 (0.17 to 0.55)	0.42
Hand-washing at 12 weeks	154	4.12 (1.10)	282	4.45 (0.82)	0.33 (0.13 to 0.53)	0.34
Intentions at baseline	87	4.93 (1.67)	310	5.23 (1.57)	0.30 (–0.09 to 0.70)	0.19
Intentions at 4 weeks	142	4.96 (1.71)	270	6.13 (1.18)	1.17 (0.85 to 1.48)	0.80
Intentions at 12 weeks	134	4.96 (1.68)	252	6.06 (1.21)	1.11 (0.79 to 1.43)	0.75

[a] Hand-washing was scored from 1 (0–2 times a day) to 5 (≥10 times a day). Intentions were scored from 1 (strongly disagree) to 7 (strongly agree). Since these analyses were not baseline adjusted, sample size varied depending on response rates at follow-up. Baseline group comparisons were not significant. All group comparisons at 4 weeks and 12 weeks were significant at $P < .001$.

[b] Confidence interval.

Longitudinal mixed-effects regression models (see Table 4) confirmed that hand-washing intentions increased from baseline to 4 weeks to a greater extent in the intervention than in the control group (time × group interaction $F_{1,375.4} = 11.71$, $P = .001$). There was also greater improvement in the theory of planned behavior cognitions in the intervention than in the control group, chiefly due to improvement in attitude in the intervention group ($F_{1,382.2} = 14.91$, $P < .001$); the effect of the intervention on subjective norm did not reach significance ($F_{1,357.9} = 2.23$, $P = .14$) and group differences in change in

XSL•FO
RenderX

perceived behavioral control were negligible ($F_{1,360.8}$ = 0.99, P = .32) (see Table 4).

Table 4. Change in theory of planned behavior cognitions from baseline to 4 weeks in the control and intervention groups[a]

Variable	Baseline, mean (SD)		4-week follow-up, mean (SD)	
	Control	Intervention	Control	Intervention
Intentions	4.93 (1.67)	5.23 (1.57)	5.05 (1.68)	6.00 (1.23)
Attitude	5.71 (1.28)	5.73 (1.97)	5.85 (1.11)	6.28 (0.78)
Subjective norm	4.99 (1.77)	5.15 (1.60)	5.27 (1.62)	5.66 (1.31)
Perceived behavioral control	6.11 (1.50)	6.21 (1.35)	6.47 (0.81)	6.45 (1.09)

[a] These analyses were carried out only in those who completed measures of baseline intentions (control n = 87; intervention n = 310). All constructs were scored from 1 (strongly disagree) to 7 (strongly agree).

Mediation of Effects on Behavior by Cognitions

As intentions and attitudes (but not subjective norms and perceived behavioral control) were changed by the intervention, we used mediation models to test whether the intervention effect might be mediated by changes in intentions or attitudes. Results showed significant positive indirect effects of the intervention on change in hand-washing via intentions (coefficient = .15, 95% CI, .08–.26) as well as attitudes (coefficient = .16, 95% CI, .09–.26). The direct effect of the intervention on change in hand-washing dropped to nonsignificance when cognitions were included in the models, in both cases.

Effects of Moderator Variables

At baseline, female gender was associated with higher levels of hand-washing (r = .34, P < .001) and intentions (r = .36, P < .001). There were no associations between age and hand-washing frequency (r = .02, P = .69) or intentions (r = –.01, P = .82). Greater socioeconomic deprivation was associated with slightly higher levels of hand-washing frequency (r = .12, P = .02) and intentions (r = .12, P = .01). Greater perceived risk was also associated with higher levels of hand-washing frequency (r = .25, P < .001) and intentions (r = .37, P < .001).

We then examined whether significant baseline predictors of hand-washing frequency and intentions moderated the effectiveness of the intervention. MANOVA revealed a main effect of gender on hand-washing frequency and intentions ($F_{2,407}$ = 12.61, P < .001; partial $\eta 2$ = .058) but no interaction with intervention group ($F_{2,407}$ = 0.30, P = .74; partial $\eta 2$ = .001). There was also a main effect of perceived risk on hand-washing frequency and intentions ($F_{2,331}$ = 14.31, P < .001; partial $\eta 2$ = .080) but no interaction with intervention group ($F_{2,331}$ = 0.69, P = .502; partial $\eta 2$ = .004). There was no effect of socioeconomic status on hand-washing frequency and intentions ($F_{2,407}$ = 0.67, P = .51; partial $\eta 2$ = .003) and no interaction with intervention group ($F_{2,407}$ = 0.35, P = .70; partial $\eta 2$ = .002).

Although the study was not powered to test for differences between subgroups, inspection of Table 5 and Table 6 shows a trend toward higher hand-washing rates and intentions in the intervention group in both men and women, those of higher and lower socioeconomic status, those with higher and lower levels of perceived risk, and those whose level of hand-washing at baseline was less than that recommended (see Table 5 and Table 6). There was an interaction between intervention group and baseline hand-washing rates for both hand-washing frequency ($F_{1,358}$ = 11.95, P = .001, partial $\eta 2$ = .032) and intentions ($F_{1,358}$ = 11.95, P = .001, partial $\eta 2$ = .032), confirming that improvement as a result of the intervention was greater in those with lower hand-washing levels. This was due partly to ceiling effects, since none of those already reporting hand-washing at the recommended rate at baseline could improve on that measure (although some could on the hand-washing intentions measure).

XSL•FO
RenderX

Table 5. Moderator effects on hand-washing frequency in the intervention and control groups at 4-week follow-up

Variable	Control group		Intervention group	
	n	Mean (SD)	n	Mean (SD)
Gender				
Male	53	3.77 (1.03)	101	4.10 (0.10)
Female	104	4.17 (1.01)	184	4.57 (0.73)
Socioeconomic status				
Lower deprivation	111	3.99 (1.07)	215	4.39(0.86)
Higher deprivation	46	4.15 (0.92)	70	4.43 (0.86)
Perceived risk				
Lower risk	35	3.77 (1.14)	93	4.10 (1.02)
Higher risk	44	4.32 (0.91)	185	4.58 (0.69)
Baseline hand-washing				
Lower rate	42	3.40 (0.96)	146	4.08 (0.95)
Higher rate	39	4.79 (0.52)	135	4.79 (0.51)

Table 6. Moderator effects on hand-washing intentions in the intervention and control groups at 4-week follow-up

Variable	Control group		Intervention group	
	n	Mean (SD)	n	Mean (SD)
Gender				
Male	50	3.77 (1.03)	92	5.01 (1.41)
Female	92	4.17 (1.01)	178	4.57 (0.73)
Socioeconomic status				
Lower deprivation	101	4.94 (1.69)	204	6.06 (1.19)
Higher deprivation	41	5.02 (1.78)	66	6.34 (1.12)
Perceived risk				
Lower risk	32	4.67 (1.71)	88	5.72 (1.43)
Higher risk	41	5.63 (1.25)	175	6.34 (0.92)
Baseline hand-washing				
Lower rate	38	4.53 (1.66)	136	5.81 (1.40)
Higher rate	37	5.89 (1.06)	130	6.47 (0.79)

Discussion

Participants given access to the Web-based intervention had higher levels of reported hand-washing frequency and intentions for frequent hand-washing in the future than those in the control group (with or without baseline measurement). This higher level of hand-washing was maintained at 12 weeks, as predicted by our primary hypotheses. These findings provide encouraging evidence that hygienic behavior may be effectively promoted by a theory-based online intervention. The medium effect sizes for reported behavior that we observed were larger than the average for Web-based interventions [54] and similar to other online interventions based on the theory of planned behavior [55]. We predicted and observed relatively modest changes in hand-washing, which is a largely habitual behavior, but these changes would nevertheless be sufficient to be valuable if replicated across much of the population. At the time of this study, participants in both groups had been exposed to considerable media and government coverage of the need for hand hygiene during the pandemic, suggesting that our intervention could usefully add to existing public health campaigns.

Moderator analyses did not reveal any significant differences in the effectiveness of the intervention for those of high and low socioeconomic status, men and women, and those with higher and lower levels of perceived risk of infection. These analyses are important in terms of establishing the suitability of the intervention for rolling out to the general population [19], and although the study was not powered to detect subgroup differences, it is reassuring that we observed a trend toward higher hand-washing rates in all the intervention subgroups. There was no evidence that socioeconomic status had a negative

XSL•FO
RenderX

impact on hand-washing, but our intervention was unable to eliminate differences in hand-washing rates associated with gender and perceived risk of infection; additional efforts may be needed to elevate hand-washing rates among men. In the event of a serious pandemic it is likely that both perceived risk and motivation to wash their hands will increase throughout the population [12,14,25].

Our planned examination of whether the intervention changed theory of planned behavior cognitions revealed substantial effects on intentions and attitudes. Although mediation model tests cannot prove causation, the findings of the mediation analyses indicated that the data were consistent with a mediation model where attitudes and intentions mediated the effects of the intervention on behavior. However, we observed no change in subjective norms or perceived behavioral control. Perceived behavioral control was already high at baseline and so a ceiling effect likely limited the potential for the intervention to increase it further. However, there was scope for improvement in subjective norms, and since social norms are an important influence on hand-washing [56], these findings suggest it might be advisable to supplement our intervention with more effective methods of changing the perceived social desirability of hand-washing.

A major limitation of our study is that only self-reported hand-washing could be assessed, which is likely to overestimate actual levels of hand-washing [15]. There were some indications that self-reports did not simply reflect socially desirable responding: higher rates of hand-washing were associated with objective measures of intervention use, and hand-washing with gel was related to objective measures of collecting hand gel. Moreover, although it seemed likely that our self-selected sample would have had above-average levels of motivation to wash their hands, reported hand-washing rates at baseline were actually slightly below UK rates reported during the pandemic [14]. The problem remains that self-report cannot be considered to provide a definitive test of whether behavior actually changed, but observation of hand-washing within the home in large samples is intrusive and impractical. However, the aim of this study was to estimate the behavioral effects of the intervention in preparation for a large trial of intervention effects on actual infection rates. For this purpose, it was essential to show that the intervention could at least influence self-reported intentions and behavior, as these can be considered a necessary (though not sufficient) precursor of reductions in infection transmission. The large study of infection rates will then allow us to perform a more definitive test of whether any reduction in infection rates achieved by this intervention is mediated by self-reported hand-washing.

A second major limitation was that our uptake rate was less than 1 in 10, and our sample overrepresented affluent, middle-aged women. This profile is typical of those who engage with Web-based health promotion [57] and suggests that it may be necessary to supplement Web-based interventions in order to reach all segments of the population; in particular, future research should establish the most effective interventions for reaching older people and ethnic minority groups, who are typically the worst affected in pandemics. Nonetheless, the moderator analyses provided some reassurance that the intervention should prove suitable for those socially deprived people who do access Web-based health interventions, and could provide a cost-effective means of reaching much of the population quickly in a pandemic.

A strength of this study is that it pragmatically assessed the effectiveness of the intervention, by calculating the effect size when compared with a control group without baseline assessment, but it also examined the efficacy of the intervention, by calculating the effect size when compared with a control group with baseline assessment. It was valuable to demonstrate that the intervention was successful when evaluated in both these ways, but this exploratory trial was not powered to make subgroup comparisons; it would be useful in future research to specifically test whether the somewhat lower effect size observed when the control group had received baseline assessment was indeed due to the effects of completing the baseline assessments.

In conclusion, this study provided the first demonstration of the potential value of a theory-based online intervention to promote behavior intended to reduce or slow the transmission of respiratory infection. An advantage for the pandemic context is that it is feasible and inexpensive to rapidly make available an intervention of this kind to a wide population, thus preserving resources for targeting groups that may require different types of intervention. For example, since this intervention was fully automated, it could be easily disseminated by links to frequently accessed health care websites and by advertising the website in government media campaigns providing information about coping with seasonal or pandemic influenza. However, further research is first required to determine the effects of the intervention on actual infection rates.

Acknowledgments

We are grateful to Adrian Osmond and Sarah Williams for assistance in developing the Web-based intervention. The Web-based materials used for this study were created using the LifeGuide system developed as part of the national Digital Social Science program funded by the UK Economic and Social Research Council, in collaboration with the LifeGuide team. This work was supported by the UK Medical Research Council (grant number GO700459). PZ Cussons (UK) Ltd provided the hand gel free of charge. The funders had no involvement in the design, analysis, or reporting of the study. The PRIMIT team also includes R Hobbs and B Carmen.

Conflicts of Interest

None declared

XSL•FO
RenderX

Authors' Contributions

PL and LY conceived, designed and supervised the study, SM developed the intervention with LY and managed the study, WS carried out the mediator and moderator analyses, LY carried out all other analyses and drafted the manuscript, and all authors contributed to revising and agreeing the final manuscript.

Multimedia Appendix 1

Sample screen shots from PRIMIT intervention.

[PPT File (Microsoft PowerPoint Presentation), 3MB - jmir_v13i4e107_app1.ppt]

Multimedia Appendix 2

Further details of intervention development and content.

[PDF file (Adobe PDF File), 92 KB - jmir_v13i4e107_app2.pdf]

Multimedia Appendix 3

Patient information leaflet.

[PDF file (Adobe PDF File), 117 KB - jmir_v13i4e107_app3.pdf]

Multimedia Appendix 4

PRIMIT protocol.

[PDF file (Adobe PDF File), 291 KB - jmir_v13i4e107_app4.pdf]

References

1. Brug J, Aro AR, Richardus JH. Risk perceptions and behaviour: towards pandemic control of emerging infectious diseases : international research on risk perception in the control of emerging infectious diseases. Int J Behav Med 2009 Mar;16(1):3-6. [doi: 10.1007/s12529-008-9000-x] [Medline: 19127440]
2. Seale H, Leask J, Po K, MacIntyre CR. “Will they just pack up and leave?” – attitudes and intended behaviour of hospital health care workers during an influenza pandemic. BMC Health Serv Res 2009;9:30 [FREE Full text] [doi: 10.1186/1472-6963-9-30] [Medline: 19216792]
3. Meier CR, Jick SS, Derby LE, Vasilakis C, Jick H. Acute respiratory-tract infections and risk of first-time acute myocardial infarction. Lancet 1998 May 16;351(9114):1467-1471. [Medline: 9605802]
4. Woodhead MA, Macfarlane JT, McCracken JS, Rose DH, Finch RG. Prospective study of the aetiology and outcome of pneumonia in the community. Lancet 1987 Mar 21;1(8534):671-674. [Medline: 2882091]
5. Brankston G, Gitterman L, Hirji Z, Lemieux C, Gardam M. Transmission of influenza A in human beings. Lancet Infect Dis 2007 Apr;7(4):257-265. [doi: 10.1016/S1473-3099(07)70029-4] [Medline: 17376383]
6. Cowling BJ, Chan KH, Fang VJ, Cheng CK, Fung RO, Wai W, et al. Facemasks and hand hygiene to prevent influenza transmission in households: a cluster randomized trial. Ann Intern Med 2009 Oct 6;151(7):437-446 [FREE Full text] [Medline: 19652172]
7. Jefferson T, Foxlee R, Del Mar C, Dooley L, Ferroni E, Hewak B, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses: systematic review. BMJ 2008 Jan 12;336(7635):77-80 [FREE Full text] [doi: 10.1136/bmj.39393.510347.BE] [Medline: 18042961]
8. Rabie T, Curtis V. Handwashing and risk of respiratory infections: a quantitative systematic review. Trop Med Int Health 2006 Mar;11(3):258-267 [FREE Full text] [doi: 10.1111/j.1365-3156.2006.01568.x] [Medline: 16553905]
9. Sandora TJ, Taveras EM, Shih MC, Resnick EA, Lee GM, Ross-Degnan D, et al. A randomized, controlled trial of a multifaceted intervention including alcohol-based hand sanitizer and hand-hygiene education to reduce illness transmission in the home. Pediatrics 2005 Sep;116(3):587-594 [FREE Full text] [doi: 10.1542/peds.2005-0199] [Medline: 16140697]
10. Bell DM, World Health Organization Writing Group. Non-pharmaceutical interventions for pandemic influenza, national and community measures. Emerg Infect Dis 2006 Jan;12(1):88-94 [FREE Full text] [Medline: 16494723]
11. Caress AL, Duxbury P, Woodcock A, Luker KA, Ward D, Campbell M, et al. Exploring the needs, concerns and behaviours of people with existing respiratory conditions in relation to the H1N1 'swine influenza' pandemic: a multicentre survey and qualitative study. Health Technol Assess 2010 Jul;14(34):1-108 [FREE Full text] [doi: 10.3310/hta14340-01] [Medline: 20630122]
12. Lau JT, Kim JH, Tsui HY, Griffiths S. Anticipated and current preventive behaviors in response to an anticipated human-to-human H5N1 epidemic in the Hong Kong Chinese general population. BMC Infect Dis 2007 Mar;7:18 [FREE Full text] [doi: 10.1186/1471-2334-7-18] [Medline: 17359545]

XSL•FO
RenderX

13. Lau JT, Yang X, Tsui HY, Pang E. SARS related preventive and risk behaviours practised by Hong Kong-mainland China cross border travellers during the outbreak of the SARS epidemic in Hong Kong. J Epidemiol Community Health 2004 Dec;58(12):988-996 [FREE Full text] [doi: 10.1136/jech.2003.017483] [Medline: 15547057]
14. Rubin GJ, Amlôt R, Page L, Wessely S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. BMJ 2009 Jul;339:b2651 [FREE Full text] [Medline: 19574308]
15. Judah G, Aunger R, Schmidt WP, Michie S, Granger S, Curtis V. Experimental pretesting of hand-washing interventions in a natural setting. Am J Public Health 2009 Oct;99 Suppl 2:S405-S411 [FREE Full text] [doi: 10.2105/AJPH.2009.164160] [Medline: 19797755]
16. Janssen AP, Tardif RR, Landry SR, Warner JE. 'Why tell me now?' the public and healthcare providers weigh in on pandemic influenza messages. J Public Health Manag Pract 2006 Aug;12(4):388-394. [Medline: 16775537]
17. Bourgeois FT, Simons WW, Olson K, Brownstein JS, Mandl KD. Evaluation of influenza prevention in the workplace using a personally controlled health record: randomized controlled trial. J Med Internet Res 2008 Mar;10(1):e5 [FREE Full text] [doi: 10.2196/jmir.984] [Medline: 18343794]
18. Glasgow RE, Klesges LM, Dzewaltowski DA, Bull SS, Estabrooks P. The future of health behavior change research: what is needed to improve translation of research into health promotion practice? Ann Behav Med 2004 Feb;27(1):3-12. [doi: 10.1207/s15324796abm2701_2] [Medline: 14979858]
19. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res 2009 Apr;11(2):e9 [FREE Full text] [doi: 10.2196/jmir.1052] [Medline: 19403465]
20. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
21. Blendon RJ, Koonin LM, Benson JM, Cetron MS, Pollard WE, Mitchell EW, et al. Public response to community mitigation measures for pandemic influenza. Emerg Infect Dis 2008 May;14(5):778-786 [FREE Full text] [Medline: 18439361]
22. Garoon JP, Duggan PS. Discourses of disease, discourses of disadvantage: a critical analysis of National Pandemic Influenza Preparedness Plans. Soc Sci Med 2008 Oct;67(7):1133-1142. [doi: 10.1016/j.socscimed.2008.06.020] [Medline: 18656294]
23. Kumar A, Zarychanski R, Pinto R, Cook DJ, Marshall J, Lacroix J, Canadian Critical Care Trials Group H1N1 Collaborative. Critically ill patients with 2009 influenza A(H1N1) infection in Canada. JAMA 2009 Nov 4;302(17):1872-1879. [doi: 10.1001/jama.2009.1496] [Medline: 19822627]
24. Nguyen-Van-Tam JS, Openshaw PJ, Hashim A, Gadd EM, Lim WS, Semple MG, Influenza Clinical Information Network (FLU-CIN). Risk factors for hospitalisation and poor outcome with pandemic A/H1N1 influenza: United Kingdom first wave (May-September 2009). Thorax 2010 Jul;65(7):645-651 [FREE Full text] [doi: 10.1136/thx.2010.135210] [Medline: 20627925]
25. Lau JT, Yang X, Tsui H, Kim JH. Monitoring community responses to the SARS epidemic in Hong Kong: from day 10 to day 62. J Epidemiol Community Health 2003 Nov;57(11):864-870 [FREE Full text] [Medline: 14600111]
26. Leung GM, Ho LM, Chan SK, Ho SY, Bacon-Shone J, Choy RY, et al. Longitudinal assessment of community psychobehavioral responses during and after the 2003 outbreak of severe acute respiratory syndrome in Hong Kong. Clin Infect Dis 2005 Jun 15;40(12):1713-1720 [FREE Full text] [doi: 10.1086/429923] [Medline: 15909256]
27. 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]
28. Green LW, Kreuter MW. Health Promotion Planning: An Educational and Ecological Approach. Mountain View, CA: Mayfield; 1999.
29. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
30. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007 May;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
31. Yardley L, Miller S, Teasdale E, Little P, Primit Team. Using mixed methods to design a web-based behavioural intervention to reduce transmission of colds and flu. J Health Psychol 2011 Mar;16(2):353-364. [doi: 10.1177/1359105310377538] [Medline: 20929941]
32. Miller S, Yardley L, Little P. Development of an intervention to reduce transmission of respiratory infections and pandemic flu: Measuring and predicting hand-washing intentions. Psychol Health Med 2011 May 17:1-23. [doi: 10.1080/13548506.2011.564188] [Medline: 21644184]
33. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
34. Godin G, Kok G. The theory of planned behavior: a review of its applications to health-related behaviors. Am J Health Promot 1996 Dec;11(2):87-98. [Medline: 10163601]
35. Sutton S. Using social cognition models to develop health behaviour interventions: problems and assumptions. In: Rutter D, Quine L, editors. Changing Health Behaviour: Intervention and Research With Social Cognition Models. Buckingham: Open University Press; 2002:193-208.

XSL•FO
RenderX

36. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
37. Rippetoe PA, Rogers RW. Effects of components of protection-motivation theory on adaptive and maladaptive coping with a health threat. J Pers Soc Psychol 1987 Mar;52(3):596-604. [Medline: 3572727]
38. Leventhal HA, Brissette I, Leventhal EA. The common-sense model of self-regulation of health and illness. In: Cameron LD, Leventhal HA, editors. The Self-regulation of Health and Illness Behaviour. London: Routledge; 2003:42-65.
39. Hobbis IC, Sutton S. Are techniques used in cognitive behaviour therapy applicable to behaviour change interventions based on the theory of planned behaviour? J Health Psychol 2005 Jan;10(1):7-18; discussion 37. [doi: 10.1177/1359105305048549] [Medline: 15576494]
40. Oettingen G, Gollwitzer PM. Goal setting and goal striving. In: Brewer MB, Hewstone M, editors. Emotion and Motivation (Perspecitves on Social Psychology). Oxford: Blackwell Publishers; 2004:165-183.
41. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
42. Roland M, Torgerson DJ. What are pragmatic trials? BMJ 1998 Jan 24;316(7127):285 [FREE Full text] [Medline: 9472515]
43. Godin G, Sheeran P, Conner M, Delage G, Germain M, Bélanger-Gravel A, et al. Which survey questions change behavior? Randomized controlled trial of mere measurement interventions. Health Psychol 2010 Nov;29(6):636-644. [doi: 10.1037/a0021131] [Medline: 20939639]
44. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries NK, Brug J. An exploration of factors related to dissemination of and exposure to internet-delivered behavior change interventions aimed at adults: a Delphi study approach. J Med Internet Res 2008 Apr;10(2):e10 [FREE Full text] [doi: 10.2196/jmir.956] [Medline: 18417443]
45. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
46. Brewer NT, Cuite CL, Herrington JE, Weinstein ND. Risk compensation and vaccination: can getting vaccinated cause people to engage in risky behaviors? Ann Behav Med 2007 Aug;34(1):95-99. [doi: 10.1080/08836610701495482] [Medline: 17688401]
47. Ajzen I. Nature and operation of attitudes. Annu Rev Psychol 2001 Feb;52:27-58. [doi: 10.1146/annurev.psych.52.1.27] [Medline: 11148298]
48. Rhodes RE, Blanchard CM, Matheson DH. A multicomponent model of the theory of planned behaviour. Br J Health Psychol 2006 Feb;11(Pt 1):119-137. [doi: 10.1348/135910705X52633] [Medline: 16480559]
49. Brewer NT, Chapman GB, Gibbons FX, Gerrard M, McCaul KD, Weinstein ND. Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. Health Psychol 2007 Mar;26(2):136-145. [doi: 10.1037/0278-6133.26.2.136] [Medline: 17385964]
50. Leppin A, Aro AR. Risk perceptions related to SARS and avian influenza: theoretical foundations of current empirical research. Int J Behav Med 2009 Feb;16(1):7-29. [doi: 10.1007/s12529-008-9002-8] [Medline: 19214752]
51. Mimas. Census Dissemination Unit, Mimas, University of Manchester. 2011. GeoConvert URL: http://geoconvert.mimas.ac.uk/ [accessed 2011-10-06] [WebCite Cache ID 62EMDvXIX]
52. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008 Nov;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
53. Mackinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: distribution of the product and resampling methods. Multivariate Behav Res 2004 Jan 1;39(1):99. [doi: 10.1207/s15327906mbr3901_4] [Medline: 20157642]
54. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011 Feb;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
55. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
56. Curtis VA, Danquah LO, Aunger RV. Planned, motivated and habitual hygiene behaviour: an eleven country review. Health Educ Res 2009 Aug;24(4):655-673 [FREE Full text] [doi: 10.1093/her/cyp002] [Medline: 19286894]
57. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an Internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595 [FREE Full text] [doi: 10.1093/her/cyp063] [Medline: 19897515]

Abbreviations

CI: confidence interval
MANOVA: multivariate analysis of variance

XSL•FO
RenderX

PRIMIT: PRimary care trial of a website based Infection control intervention to Modify Influenza-like illness and respiratory infection Transmission
SARS: severe acute respiratory syndrome

Edited by G Eysenbach; submitted 11.10.11; peer-reviewed by F Bourgeois, V Curtis, A Card; comments to author 29.10.11; revised version received 09.11.11; accepted 17.11.11; published 09.12.11.

Please cite as:
Yardley L, Miller S, Schlotz W, Little P
Evaluation of a Web-Based Intervention to Promote Hand Hygiene: Exploratory Randomized Controlled Trial
J Med Internet Res 2011;13(4):e107
URL: http://www.jmir.org/2011/4/e107/
doi:10.2196/jmir.1963
PMID:22155673

©Lucy Yardley, Sascha Miller, Wolff Schlotz, Paul Little. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 09.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploratory Study of Web-Based Planning and Mobile Text Reminders in an Overweight Population

Anastasia Soureti[1,2], BSC(Hons), MSc; Peter Murray[1], CStat, Mphil; Mark Cobain[1], PhD; Mai Chinapaw[2], PhD; Willem van Mechelen[2], MD; Robert Hurling[1], PhD

[1]Unilever Discover, Bedforshire, United Kingdom

[2]Department of Public and Occupational Health and EMGO Institute for Health and Care Research, VU University Medical Centre, Amsterdam, Netherlands

Corresponding Author:
Anastasia Soureti, BSC(Hons), MSc
Unilever Discover
Colworth Science Park
Bedforshire, MK44 1LQ
United Kingdom
Phone: 44 44 1234 248189
Fax: 44 44 1234 24 8010
Email: natasha.soureti@unilever.com

Abstract

Background: Forming specific health plans can help translate good intentions into action. Mobile text reminders can further enhance the effects of planning on behavior.

Objective: Our aim was to explore the combined impact of a Web-based, fully automated planning tool and mobile text reminders on intention to change saturated fat intake, self-reported saturated fat intake, and portion size changes over 4 weeks.

Methods: Of 1013 men and women recruited online, 858 were randomly allocated to 1 of 3 conditions: a planning tool (PT), combined planning tool and text reminders (PTT), and a control group. All outcome measures were assessed by online self-reports. Analysis of covariance was used to analyze the data.

Results: Participants allocated to the PT ($mean_{saturatedfat}$ 3.6, $mean_{copingplanning}$ 3) and PTT ($mean_{saturatedfat}$ 3.5, $mean_{copingplanning}$ 3.1) reported a lower consumption of high-fat foods ($F_{2,571}$ = 4.74, P = .009) and higher levels of coping planning ($F_{2,571}$ = 7.22, P < .001) than the control group ($mean_{saturatedfat}$ 3.9, $mean_{copingplanning}$ 2.8). Participants in the PTT condition also reported smaller portion sizes of high-fat foods (mean 2.8; $F_{2,569}$ = 4.12, P = .0) than the control group ($mean_{portions}$ 3.1). The reduction in portion size was driven primarily by the male participants in the PTT (P = .003). We found no significant group differences in terms of percentage saturated fat intake, intentions, action planning, self-efficacy, or feedback on the intervention.

Conclusions: These findings support the use of Web-based tools and mobile technologies to change dietary behavior. The combination of a fully automated Web-based planning tool with mobile text reminders led to lower self-reported consumption of high-fat foods and greater reductions in portion sizes than in a control condition.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 61819220; http://www.controlled-trials.com/ISRCTN61819220 (Archived by WebCite at http://www.webcitation.org/63YiSy6R8)

(J Med Internet Res 2011;13(4):e118) doi:10.2196/jmir.1773

KEYWORDS

Implementation intentions; mobile text reminders; saturated fat

Introduction

A healthy diet low in saturated fat is a popular recommendation in helping overweight and obese individuals eat healthier and reduce their health risk factors [1]. Since sustained change is very difficult, it is important to identify approaches that help people maintain new healthy behaviors once initiated. We have used the Health Action Process Approach model [2-4] to explore behavioral change, as it provides a theoretical framework on how to maintain the translation of intentions into action.

XSL•FO
RenderX

Since good intentions are not always translated into action, the emphasis was on using specific plans also known as implementation intentions. A meta-analysis of 94 studies [5] showed that implementation intentions had a positive effect of medium to large magnitude on goal achievement. Several different formats have been used in the past [6-8]. We used the *if...then* format in the present study. The *if* statements are purported to increase the accessibility of critical situations to a person, while the *then* component of the plan creates a stronger link between the situational cue and the goal-directed response. Promising results of this format have been reported in several health-related behaviors [9-12].

Most of the studies on implementation intentions have used face-to-face communication. These studies showed a reliable effect on diet, physical activity, and smoking cessation [5,8-11], whereas evidence from the two studies on the effects of Web-based interventions was mixed [12,13]. In a study conducted in an occupational setting, use of Web-based implementation intentions backfired, such that participants who did not form an implementation intention exercised significantly more than participants who formed an implementation intention [13]. In a recent dietary intervention, participants allocated to a Web-based implementation intention condition reported a reduction in their self-reported saturated fat intake [12].

Reminding people of their plans could enhance the impact of implementation intentions on behavior [14,15]. Delivering strategies via mobile phone technology is particularly appealing because of the widespread use of mobiles phones in the United Kingdom (where this study is being conducted), Europe, and the United States [16,17]. Penetration in Europe has surpassed the 100% mark [18]. Although the application of mobile and text-based technology for behavior change is in its infancy, there is some supportive evidence for physical activity, exercise behavior, and smoking cessation [19-24]. In one of these studies aimed at improving exercise behavior [19], participants were randomly allocated to 1 of 5 conditions (implementation intentions and text reminders, implementation intentions, text reminders, and 1 of 2 control groups). In the follow-up 4 weeks later, results suggested a superiority of the combined condition in the frequency of exercise, while neither the text reminders nor the implementation intention conditions alone were effective. In a later study, the same authors discussed that pairing implementation intentions with goal reminders increased the level of brisk walking and appeared to activate other related health behaviors such as weight loss [20].

Objectives

The primary aim of the present study was to assess the effects of a fully automated planning tool and mobile text reminders on participants' reduced intake of high-fat foods and food portion sizes. We hypothesized that participants in the combined planning tool and text reminders (PTT) condition would reduce their consumption of high-fat foods and make greater changes in their portion sizes. A secondary aim of this study was to assess changes in other health-related behaviors not directly targeted by the intervention. We also investigated the intervention effects on social cognitive measures (eg, self-efficacy, planning, and intentions) and the mediating properties of planning.

Methods

Participants

Participants were recruited by a recruitment agency via an online questionnaire that screened them for eligibility. The self-reported eligibility criteria were age (30–60 years), weight (body mass index [BMI] >25 kg/m^2), not having a diagnosis of a heart condition or cancer, not being pregnant, and being a mobile phone user, happy to receive text reminders, computer savvy, and motivated to change their dietary patterns. Only motivated individuals were included, since they are closer to enacting their behavior and are known to make better use of a planning regimen [25,26]. We also chose overweight participants because, though at no immediate risk of disease, they would be more likely than normal-weight individuals to benefit from planning and saturated fat reductions. We selected individuals in the age range between 30 and 60 years because this is the age when they are starting to get more interested in their long-term health. Allocation of the participants to the 3 conditions at the intervention stage was stratified by age group (30–45 years or 46–60 years) and gender.

Design and Procedures

This study was conducted between January and March 2010 and was registered retrospectively (ISRCTN61819220). This was an exploratory randomized controlled, between-groups study with no participant–experimenter contact. Participants were given online instructions and completed each week's session from the convenience of their home computer. At week 1, participants were recruited by an online agency, signed an online consent form [27] (Multimedia Appendix 1), and were then randomly allocated into 1 of 3 conditions using a computer-generated list of random numbers: (1) control group, (2) planning tool (PT), and (3) PTT. Participants in the PT condition logged in to a Web-based, fully automated program, where they were prompted to identify a list of situations to change their saturated fat intake. The program then guided them to match these situations with a list of behaviors. They completed the session once and were not able to revisit the website to make any changes, such as to create more plans. Participants in the PTT condition first used the planning tool and then were offered text reminders of their plans. All 3 groups received educational information on the importance of a healthy diet low in saturated fat, and the association between high cholesterol and being overweight was highlighted (Multimedia Appendix 1). The term bad fats was used to refer to saturated fat. At the end of week 1, all participants completed an online questionnaire on their current saturated fat intake, maintenance self-efficacy, and intentions to change their dietary intake. A coupon for a cholesterol-lowering product, with £0.50 off the face value, was mailed to all participants as a reward for successfully completing baseline assessments.

Participants revisited the website 4 weeks later (week 5) and completed a follow-up online questionnaire on maintenance and recovery self-efficacy, action and coping planning, saturated

XSL•FO
RenderX

fat intake, and portion size changes. Participants received £15 upon study completion and were entered in a prize draw for vouchers (£200 value).

Interventions

The Planning Tool

Participants who received the planning tool selected from a list of 13 situations, in which they were tempted to eat unhealthily and then chose an approach to change their behavior from a list of 13 solutions. The solutions were based on constructs from the Processes of Change Model (eg, counterconditioning, stimulus control, and helpful relationships) [28]. Several nutritionally based behaviors were also included from an accredited site [29] after review by an expert nutritionist. The list of situations consisted of both situational cues (eg, having lunch) focusing on the *when and where* and motivational cues (eg, feeling bored) linked to the reasons (*why*) for performing a specific behavior [30]. Motivational cues were divided into 3 main situations: (1) experiencing positive affect, (2) experiencing negative affect, and (3) being faced with cravings [31,32]. The situations were translated into *if* statements (eg, "If I'm having breakfast") and the solutions were translated into *then* statements (eg, "then I will tell myself I can eat healthily").

For every situation–solution pair chosen, the program drew a line on the computer screen to visually link the two together [10]. Participants were asked to complete at least 3 situation-solution pairs. Once these pairs were chosen and saved, participants were not able to revisit the program to change them during the 4-week period. The planning tool is shown in Figure 1.

Figure 1. Web-based, fully automated planning tool for changing dietary behavior.



Text Reminders

After completing the planning session, participants in the PTT entered their mobile number and chose a time band to receive text reminders of their plans.

Control Group

At the end of the study, participants in the control group received educational information on the importance of a healthy diet low in saturated and on the association between high cholesterol and being overweight. At weeks 1 and 5, they filled out the same online questionnaires as the participants in the rest of the experimental conditions .

Outcome Measures and Statistical Considerations

Behavioral Outcomes

Behavior was assessed by the following measures.

(1) A *food frequency questionnaire* [33], which records the frequency of consumption of 63 common foods. This food frequency questionnaire has good test–retest reliability (r = .62, P < .01) [33] and validity when compared with 10-day weighed records [33-35].

(2) A *2-item scale* (r = .79, P < .001) adapted from a previous study [34]. Participants were asked to report on a 7-point Likert scale their agreement on consumption of high-saturated fat foods ("I have eaten foods high in bad fats in the last week") followed by frequency of consumption of these foods ("How often did you eat foods high in bad fats in the last week?").

(3) *Portion size changes* in the consumption of 11 items (eg, meat dishes, whole milk, bacon, ordinary cheese, chocolate, chips). These items accounted for the highest reported saturated fat intake in the food frequency questionnaire from a previous study [12]. Participants were asked to report changes in their portion sizes on a 7-point Likert scale (from a lot less to a lot

XSL•FO
RenderX

more). The items were highly correlated with each other (r = .92) and were analyzed together as a composite score.

(4) *Other health behaviors* measured by 6 items (Cronbach alpha = .778) on a 7-point Likert scale (from a lot less to a lot more). Participants were asked to report changes in other health areas, namely alcohol, use of cholesterol-lowering products, weight changes, smoking, physical activity, and eating a well-balanced diet.

Social Cognitive Outcomes

(1) *Intentions* to reduce consumption of high-fat foods were assessed on a 7-point Likert scale by 3 items (eg, "I intend to eat smaller portions of high-fat foods, replace high fat with low-fat alternatives"). Due to good reliability (r = .88), the mean of the 3 items was used in the analysis.

(2) *Maintenance and recovery self-efficacy* were modified from previous research [34,36,37]. Maintenance self-efficacy (Cronbach alpha = .79), which was measured at all study times, consisted of 3 items focusing on confidence at sustaining change in the face of difficulties. It consisted of items such as "I am certain that I could overcome difficulties when trying to eat more healthily even if...I don't see success at once, I won't get support for my first attempts." Recovery self-efficacy (Spearman r = .67) was assessed only at week 5. It consisted of 2 items exploring confidence to start eating healthily when lapses occur. Items were measured on a 4-point scale (not at all, barely true, mostly true, exactly true).

(3) *Action and coping planning* were adapted from previous research [3,4,36-39]. Action planning consisted of 2 items ("I now have my own plan regarding a) when and b) how to eat more healthily") (Spearman r = .77). Coping planning consisted of 3 items (Cronbach alpha = .84) and focused on having a plan to deal with barriers (eg, "I have a detailed plan how to avoid high-risk situations where the urge to eat unhealthy food is high").

(4) *Feedback on the intervention* was assessed at week 5 on a 7-point Likert scale (strongly disagree to strongly agree). Participants were asked to rate the intervention in terms of its personal relevance, interest, trustworthiness, credibility, feelings of enjoyment, and worry. All items were adapted from previous studies [40-42].

Statistical Considerations

This was an intention-to-treat analysis based on those participants who successfully completed week 1 assessments (n = 808). Data from the food frequency questionnaire were summarized to yield the total calorie intake per participant and the percentage of total energy intake due to saturated fat. Analysis of variance with baseline covariates was conducted to analyze the behavioral (2-item scale, food frequency questionnaire, portion sizes changes, and other health behaviors) and social cognitive measures (intentions, self-efficacy, planning, and feedback). Baseline self-reported saturated fat intake was included as a covariate for the analysis of the primary outcome measures. Other potential covariates (eg, smoking, BMI, and social economic status) were retained if significant in the analysis. For any significant condition effects, we conducted Tukey–Kramer-adjusted multiple comparisons to test for further differences between the groups. All of the above analysis was carried out using version 9.2 of SAS software (SAS Institute, Cary, NC, USA). To test the mediating properties of planning, we used the SPSS macro developed by Preacher and Hayes [43]. This program uses a bootstrapping resampling strategy to evaluate significance of the model and effects of the mediator; for this analysis, 5000 bootstrap samples were used [33-35]. In the following steps, we assessed (1) the relationship between the intervention and the 2-item scale, (2) the relationship between the intervention and planning, and 3) the relationship between the intervention and the 2-item scale, taking into account planning. The bootstrapped a*b path of planning was calculated to test the significance of the mediator. For the mediation analysis, the intervention was coded as PT and PTT = 1 and control group = 0.

Local Research Ethical Review Requirement

The study was approved by an independent research ethics committee (Colworth Research Ethics Committee) in the South of England on October 21, 2009 (Multimedia Appendix 2). All research was conducted in accordance with the Helsinki Declaration [44].

Results

Participant Baseline Characteristics

At week 1, we contacted 1013 participants, of whom 96 did not meet the study's inclusion criteria. We randomly allocated 858 into 1 of 3 study conditions, and 808 of them completed week 1. Table 1 shows participants' baseline characteristics at week 1.

At week 5, a total of 571 participants revisited the website (70.7%). The number of participants completing each week is shown in Figure 2.

XSL•FO
RenderX

Table 1. Participant baseline characteristics (week 1)

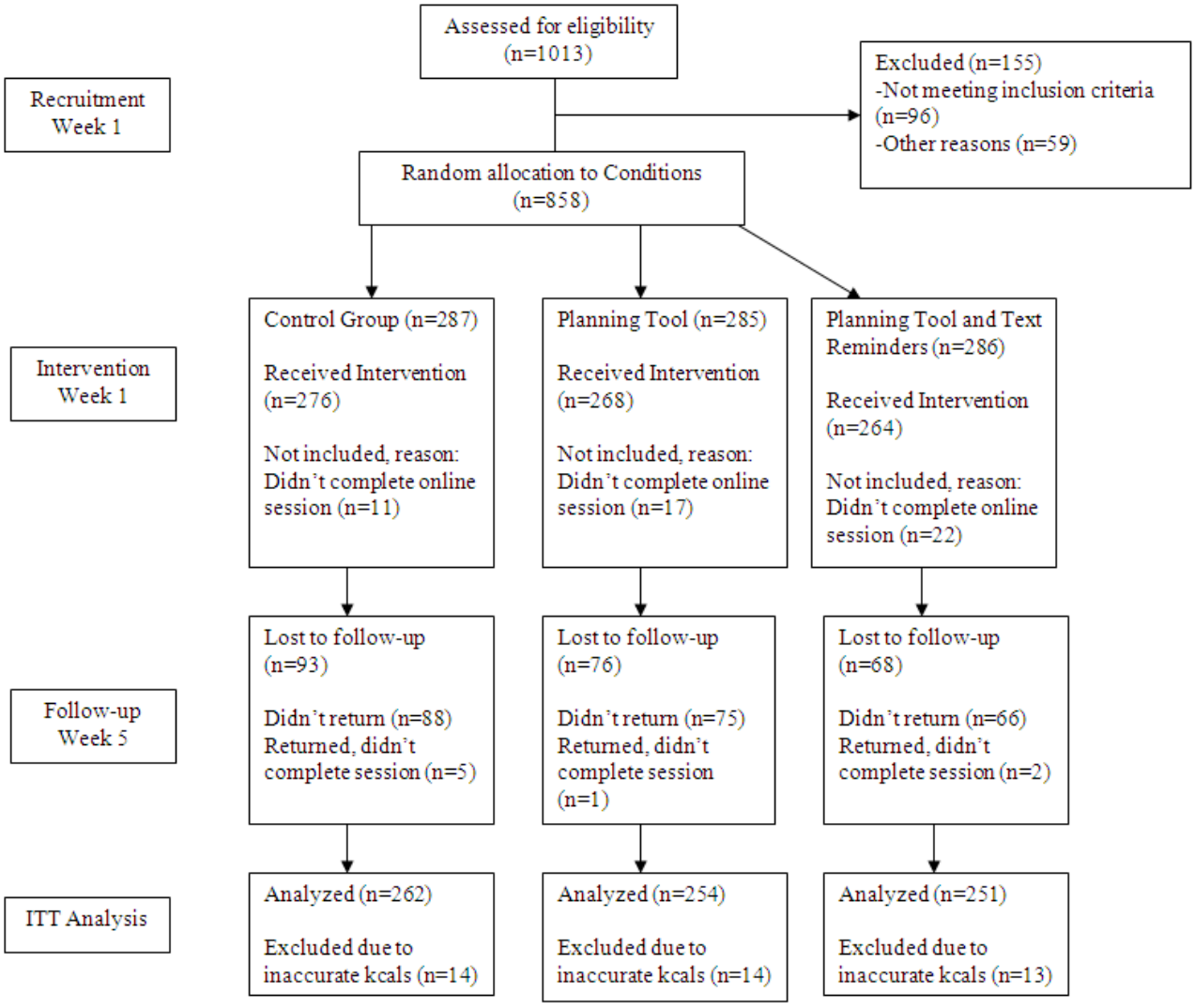
	Overall (n = 808)	$F_{2,805}$ statistic	*P* value	Control group (n = 276)	Planning tool group (n = 268)	Planning tool and text reminders group (n = 264)
Age (years), mean (SD)[a]	46.0 (8.6)	0.23	.79	45.9 (8.4)	46.2 (8.6)	45.8 (8.7)
BMI[b] (kg/m^2), mean (SD)[a]	31.7 (5.9)	1.08	.34	31.3 (5.3)	31.7 (5.4)	32.1 (7.0)
% Saturated fat, mean (SD)[a]	15 (3.1)	0.93	.39	15 (2.9)	15 (3.2)	15 (3.1)
Smokers, n (%)[a,c]	24	0.09	.95	66/276 (23)	66/268 (24)	66/264 (23)

[a] No significant differences found in participants' baseline characteristics ($P > .05$).

[b] Body mass index.

[c] χ^2 test statistic.

Figure 2. Flowchart of recruitment, intervention, and follow-up (ITT analysis = intention-to-treat analysis).



Planning Tool

All participants allocated to the PT condition were able to formulate their plans online, and 85% completed 4 situation–solution pairs (83% PT and 87% PTT). The most frequently chosen situations were "If I'm getting a snack" (25%), "If I'm feeling hungry" (21%), "If I'm in a restaurant" (20%), and "If I feel like having a chocolate" (19%). The most frequently selected solutions included action plans such as "Then I will go for fruit" (18%), "Then I will go for chicken or fish" (15%), "Then I will have a smaller portion" (11%), and "Then I will find out about a lower fat option" (9%).

Impact of the Intervention on Behavioral Outcomes

(1) The *2-item scale*: Participants in all 3 conditions reported a significant reduction in consumption of foods high in saturated

XSL•FO
RenderX

fats between baseline and follow-up (Table 2). Analysis of covariance showed a significant difference between the conditions ($F_{2,571}$ = 4.74, P = .009) with respect to self-reported saturated fat intake changes. Tukey–Kramer multiple comparisons indicated that this was due to participants in the PT and PTT conditions reporting a lower self-reported consumption of high-fat foods at week 5 than those in the control group.

(2) *Food frequency questionnaire*: Participants in all conditions reported a significant reduction in percentage saturated fat intake between baseline and follow-up (Table 2) with no significant between-group differences (P = .23).

(3) *Portion size changes*: Participants in all conditions reported a significant reduction in their portion sizes for high saturated fat foods at week 5 (Table 2). There was a significant effect of condition on portion size changes ($F_{2,569}$ = 4.12, P = .017). Tukey–Kramer-adjusted multiple comparisons indicated a significant difference between participants in the PTT and control group (P = .02). Furthermore, a significant condition-by-gender interaction was found ($F_{3,569}$ = 3.29, P = .02), with men in the PTT condition (mean 2.86, SE 0.084) reporting a greater reduction in their portion sizes than men in the control group (mean 3.27, SE 0.09; P = .003). This intervention effect was not found for women, who reported an average reduction of 2.92 (SE 0.01) in the PTT condition, 2.78 (SE 0.11) in PT, and 3.04 (SE 0.10) in the control group.

(4) *Other health behaviors*: Analysis of covariance showed that there was no significant condition effect for any of the other health behaviors apart from "I ate a well-balanced diet" ($F_{2,570}$ = 5.3, P = .005). Tukey–Kramer-adjusted multiple comparisons found a significant difference in eating a well-balanced diet between the PTT condition and the control group (P = .004).

Table 2. Saturated fat intake as measured by mean scores on a two-item scale, food frequency questionnaire, and portion size changes

	Control group (n = 276)			Planning tool (n = 268)			Planning tool and text reminders (n = 264)		
	Week 5	Week 5–1[a]	t_{275} (P value)	Week 5	Week 5–1[a]	t_{268} (P value)	Week 5	Week 5–1[a]	t_{263} (P value)
2-item scale	3.9	–1.1 (0.1)	<.001	3.6	–1.5 (0.1)	<.001	3.5	–1.2 (0.1)	<.001
Food frequency questionnaire	14.9	–0.7 (0.2)	<.001	14.5	–1.1 (0.2)	<.001	14.8	–0.8 (0.2)	<.001
Portion sizes	3.15	NA[b]	<.001	3.0	NA[b]	<.001	2.88	NA[b]	<.001

[a] Mean change and standard errors after adjusting for baseline and other covariates.

[b] Not applicable: portion size changes were assessed only at week 5, so mean change (week 5–1) was not calculated.

Impact of the Intervention on Social Cognitive Variables

There was no significant condition effect on intention to reduce consumption of high-fat foods ($F_{2,570}$ = 1.7, P = .18), maintenance self-efficacy ($F_{2,571}$ = 0.7, P = .49), or recovery self-efficacy ($F_{2,571}$ = 0.2, P = .86). There was a significant effect of condition on coping planning ($F_{2,571}$ = 7.2, P < .001) but not on action planning (P = .16). For coping planning there was a significant difference between the PPT (mean 3.1; P < .001) and control group (mean 2.8), and PT (mean 3.0; P = .02) and control group, but not between the PTT and PT. Due to the above result, only the mediating properties of coping planning were tested. There was a significant relationship between the intervention (PTT and PT vs control group) and coping planning (beta = .23; 95% CI, .11–.35) and the intervention and the 2-item scale (beta = –.35; 95% CI, –.59 to –.11). Results from bootstrapping showed a significant mediation effect of coping planning (a*b = –.16; 95% CI, –.27 to –.08). We observed partial mediation since the effect of the intervention on self-reported saturated fat intake became nonsignificant and the b value was reduced but did not become zero (beta = –.20; 95% CI, –.42 to .02). At week 5, analysis of variance showed that there were no significant between-group differences in most of the feedback on the intervention items, apart from the worry item ($F_{2,570}$ = 7.5, P < .001), with participants in the PTT condition (mean 3.21) reporting the information being less worrying than the control group (mean 3.9) (P < .001).

Discussion

We found that participants completing a Web-based, fully automated planning tool alone or in combination with mobile text reminders self-reported a greater reduction in the consumption of high-fat foods than the control group. Participants in the PTT condition also reported higher levels of coping planning. Men in the PTT condition reported a greater reduction in the portion sizes of high-fat foods. Participants in the PT group without reminders did not change portion sizes more than those in the control group but showed higher levels of coping planning than the control group. We found no significant group differences in terms of percentage saturated fat intake, intentions, self-efficacy, or feedback on the intervention.

This study provides some support for the combined effects of implementation intentions plus text message reminders in reductions of high-fat foods (measured by the 2-item scale) and portion sizes. The two previous studies that combined implementation intentions with text messages found positive effects within the arena of physical activity and exercise behavior [18,19]. If implementation intentions operate by connecting the environmental cue with a desired response, then adding text messages may further strengthen either this

XSL•FO
RenderX

connection or someone's commitment to enact his or her plan [14].

Planning on its own showed a significant difference with the control group in relation to self-reported reductions in the 2-item scale but no significant differences for any of the other behaviors (eg, portion sizes, percentage saturated fat intake, and other health behaviors). Most of the evidence on the positive effects of implementation intentions comes from offline research studies. Forming plans online might be different from using a paper-and-pencil experiment. For example, in a Web-based study planning backfired so that people who did not form implementation intentions exercised significantly more than those who formed a Web-based implementation intention [13]. It is also the case that forming plans might not be effective, on their own, in some circumstances. For example, in a recent study [19], use of planning plus text reminders led to greater increases in exercise behavior than in the control group, yet there was no difference from control for those making plans with no text reminders.

Participants in all conditions reported a significant reduction in percentage saturated fat intake from the food frequency questionnaire and the 2-item scale. The existence of sufficiently high levels of motivation at the beginning of the study or a repeat measurement effect [45] might help explain this universal change and why, contrary to our expectations, there were no significant differences between groups for the percentage saturated fat measure. However, the 2-item scale was better able to differentiate between conditions than the food frequency questionnaire measure. Similar discrepancies in findings between percentage saturated fat intake and the 2-item scale have been reported before [9,12]. Both come with limitations, which we discuss later in this section.

The effects of the combined implementation intentions and text message condition on portion size changes were driven by male participants. Research on young females and males has suggested a difference in preferred communication style and use of technologies [46, 47]. Younger females (15–19 years old) tend to prefer in-depth conversations and to write longer messages with a more complex structure, while young males tend to be oriented toward simple one-thought messages and more task-oriented conversations [48, 49]. If reliable, these gender communication preferences would explain the differentiated impact of the text message reminders of plans, which was a rather task-oriented activity. An alternative explanation is that there was more scope for men to reduce their portion sizes. This could have led to a regression to the mean effect for men, so enhancing the apparent change in portion size. However, it has previously been reported that men tend to self-report consumption of bigger portions for solid, high-energy, and high-fat foods [50], main meals, and side dishes [51]. If this were true in our study then the larger reduction in portion size may have made it easier for men to reduce their portion sizes. The measurement of portion size change at a single point is a weakness of this study. Further studies are needed to compare changes in self-reported portion sizes by including pre- and postintervention measures.

We found no effect of the combined implementation intention and text reminders condition on other health areas (eg, weight loss and physical activity). In the current study, implementation intentions were paired up with plan reminders. Other research [20] has suggested that text reminders of people's *goals* (eg, being healthier), but not plans, may activate related health behaviors (eg, dietary restriction) leading to other outcomes (eg, weight loss).

Contrary to our expectations and some previous research, there were no differences between the implementation intention conditions and the control group for social cognitive measures such as participants' maintenance and recovery self-efficacy [12]. This might be because participants, who formed plans and encountered obstacles, needed further support to maintain their healthy eating and recover from lapses. Also, our intervention was not designed to increase self-efficacy beliefs. As suggested by the Health Action Process Approach model, strong self-efficacy beliefs could be the precursors of planning and behavioral change [4]. Indeed, self-efficacy has been found to predict whether people make plans about physical activity [52] and mediate the effects of planning on weight control [53]. Future interventions could first try boosting self-efficacy beliefs (through direct mastery experience or modeling) before the planning session.

Participants in the PTT and PT conditions chose mostly action rather than coping plans at week 1. Although there was no difference between conditions in the reported use of action plans at week 5, the participants in the PTT and PT conditions at week 5 reported using more coping planning than did the control group, and coping planning partially mediated the effects of the intervention on self-reported saturated fat intake. Hence, it seems that the use of coping plans was instrumental in driving changes in behavior. Previous research has suggested that action plans are more effective at the early stages of change, while coping plans are instrumental at later stages, and that a combined action and coping planning condition is better than a condition focusing solely on action plans [37,38]. In contrast to our results, a recent study on simultaneous use of action and coping plans found that they both mediated the intervention effects on fruit and vegetable intake [54]. Further studies are needed to more systematically explore the relative effectiveness of action and coping plans.

The impact of conditions on our measures of saturated fat intake changes was inconsistent, and this could be due to the limitations of the food frequency questionnaire and the 2-item scale. Underreporting of food consumption is a recurrent challenge for food frequency questionnaires and is most pronounced among overweight and obese people [55]. Also, food frequency questionnaires were initially designed to estimate individual intake relative to a population rather than to detect small changes in individual dietary intake [56], for which they might not be sufficiently sensitive. The original food frequency questionnaire calculation did not account for individual variation in portion sizes but instead assumed the average portion of the UK population [33]. When we considered portion sizes in this study, we found a significant difference between the PTT and the control group. On the other hand, self-reported items such as the 2-item scale were designed to detect differences between

XSL•FO
RenderX

conditions in experimental studies [8,9]. However, some have claimed that reported changes are influenced by demand characteristics [57], with participants in more active conditions being more aware of study aims and so responding differently. Two previous studies counter the argument of demand characteristics by showing no difference between conditions for awareness of the study's hypothesis or feelings of obligation to comply [9,13].

The current study has several advantages. It is based on a rather rigorous statistical analysis (ie, intention-to-treat analysis) and was conducted in a real-life setting. With a few exceptions most studies so far have been laboratory based and have examined effects of behavior in student populations [58,59]. This study also focused on overweight individuals, who are most likely to benefit from dietary interventions. Our planning tool is the first fully automated, interactive system to test Web-based if...then plans in the format of an interactive volitional help sheet, with the advantage of letting people choose the more personally relevant situations. Also, text messages were deployed with some flexibility by allowing participants to choose appropriate delivery times.

Limitations of the current study include the use of self-reports and the short-term follow-up of 4 weeks. It is essential that the combined effects of text messages and implementation intentions be tested over longer periods of time (ie, 6 months and 1 year) for sustained behavior change and that more objective measures of dietary behavior change be incorporated. Future research could also benefit by using ecological momentary assessment [60] so that text reminders of plans are sent at risky occasions, where cravings for high-fat foods occur.

Conclusions

In conclusion, fully automated, Web-based planning tools and mobile technologies provide a good opportunity to promote large-scale dietary behavior change in overweight adults. In this exploratory study, a combined Web-based planning and text message condition was associated with some reductions in self-reported consumption of high saturated fat foods in comparison with the no-treatment control group.

Acknowledgments

Andrew Hinton and Steve Baker from Tessella PLC had full access to all of the data and take responsibility for the integrity of the data. Virtual Surveys PLC was in charge of the online recruitment of participants. This work was supported by Unilever Discover, Colworth House, Sharnbrook, Bedfordshire MK44 1LQ.

Conflicts of Interest

This exploratory study was funded by Unilever, where several of the authors are employees. The planning tool is a non-commercially available research vehicle.

Multimedia Appendix 1

Research protocol (in lieu of prospective trial registration authors provide the research protocol).

[PDF File (Adobe PDF File), 1MB - jmir_v13i4e118_app1.pdf]

Multimedia Appendix 2

Ethics Committee Acceptance Letter.

[PDF File (Adobe PDF File), 93KB - jmir_v13i4e118_app2.pdf]

References

1. Artinian NT, Fletcher GF, Mozaffarian D, Kris-Etherton P, Van Horn L, Lichtenstein AH, American Heart Association Prevention Committee of the Council on Cardiovascular Nursing. Interventions to promote physical activity and dietary lifestyle changes for cardiovascular risk factor reduction in adults: a scientific statement from the American Heart Association. Circulation 2010 Jul 27;122(4):406-441 [FREE Full text] [doi: 10.1161/CIR.0b013e3181e8edf1] [Medline: 20625115]
2. Schwarzer R. Self-regulatory processes in the adoption and maintenance of health behaviors. J Health Psychol 1999 Mar;4(2):115-127. [doi: 10.1177/135910539900400208] [Medline: 22021474]
3. Schwarzer R, Sniehotta FF, Lippke S, Luszczynska A, Scholz U, Schuz B, et al. Freie Universität Berlin. 2003 Nov 1. On the Assessment and Analysis of Variables in the Health Action Process Approach: Conducting an Investigation URL: http://web.fu-berlin.de/gesund/hapa_web.pdf [accessed 2011-06-27] [WebCite Cache ID 5zkvk39uc]
4. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
5. Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement: a meta-analysis of effects and processes. Adv Exp Soc Psychol 2006;38:69-111. [doi: 10.1016/S0065-2601(06)38002-1]
6. Orbell S, Sheeran P. Motivational and volitional processes in action initiation: a field study of the role of implementation intentions. J Appl Soc Psychol 2000;30(4):780-797. [doi: 10.1111/j.1559-1816.2000.tb02823.x]

XSL•FO
RenderX

7. Oettingen G, Honig, G, Gollwitzer PM. Effective self-regulation of goal attainment. Int J Educ Res 2000;33(7-8):705-732. [doi: 10.1016/S0883-0355(00)00046-X]
8. Armitage CJ. Evidence that implementation intentions reduce dietary fat intake: a randomized trial. Health Psychol 2004 May;23(3):319-323. [doi: 10.1037/0278-6133.23.3.319] [Medline: 15099174]
9. Chapman J, Armitage CJ, Norman P. Comparing implementation intention interventions in relation to young adults' intake of fruit and vegetables. Psychol Health 2009 Mar;24(3):317-332. [doi: 10.1080/08870440701864538] [Medline: 20204996]
10. Armitage CJ. A volitional help sheet to encourage smoking cessation: a randomized exploratory trial. Health Psychol 2008 Sep;27(5):557-566. [doi: 10.1037/0278-6133.27.5.557] [Medline: 18823182]
11. Prestwich A, Ayres K, Lawton R. Crossing two types of implementation intentions with a protection motivation intervention for the reduction of saturated fat intake: a randomized trial. Soc Sci Med 2008 Nov;67(10):1550-1558. [doi: 10.1016/j.socscimed.2008.07.019] [Medline: 18771834]
12. Soureti A, Murray P, Cobain MR, van Mechelen W, Hurling R. Web-based risk communication and planning in an obese population: exploratory study. J Med Internet Res 2011;13(4):e100 [FREE Full text] [doi: 10.2196/jmir.1579] [Medline: 22126827]
13. Budden JS, Sagarin BJ. Implementation intentions, occupational stress, and the exercise intention-behavior relationship. J Occup Health Psychol 2007 Oct;12(4):391-401. [doi: 10.1037/1076-8998.12.4.391] [Medline: 17953497]
14. Gollwitzer PM. Implementation intentions: strong effects of simple plans. Am Psychol 1999;54:493-503 [FREE Full text]
15. Sheeran P, Webb TL, Gollwitzer PM. The interplay between goal intentions and implementation intentions. Pers Soc Psychol Bull 2005 Jan;31(1):87-98. [doi: 10.1177/0146167204271308] [Medline: 15574664]
16. Nielsen. The Nielsen Company. 2010. 2010 Media Industry Fact Sheet URL: http://blog.nielsen.com/nielsenwire/press/nielsen-fact-sheet-2010.pdf [accessed 2011-06-30] [WebCite Cache ID 5zpeUZsGb]
17. Anonymous. BBC. 2010 Jul 09. Over 5 Billion Mobile Phone Connections Worldwide URL: http://www.bbc.co.uk/news/10569081 [accessed 2011-12-01] [WebCite Cache ID 63c3SV193]
18. Anonymous. International Telecommunication Union. 2008 Jul 15. Global ICT Developments URL: http://www.itu.int/ITU-D/ict/statistics/ict/index.html [accessed 2011-06-30] [WebCite Cache ID 5zpdufmkn]
19. Prestwich A, Perugini M, Hurling R. Can the effects of implementation intentions on exercise be enhanced using text messages? Psychol Health 2009 Jul;24(6):677-687. [doi: 10.1080/08870440802040715] [Medline: 20205020]
20. Prestwich A, Perugini M, Hurling R. Can implementation intentions and text messages promote brisk walking? A randomized trial. Health Psychol 2010 Jan;29(1):40-49. [doi: 10.1037/a0016993] [Medline: 20063934]
21. Morris ME, Kathawala Q, Leen TK, Gorenstein EE, Guilak F, Labhard M, et al. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. J Med Internet Res 2010;12(2):e10 [FREE Full text] [doi: 10.2196/jmir.1371] [Medline: 20439251]
22. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007 Apr;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
23. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin RB, et al. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. Tob Control 2005 Aug;14(4):255-261 [FREE Full text] [doi: 10.1136/tc.2005.011577] [Medline: 16046689]
24. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
25. de Nooijer J, de Vet E, Brug J, de Vries NK. Do implementation intentions help to turn good intentions into higher fruit intakes? J Nutr Educ Behav 2006;38(1):25-29. [doi: 10.1016/j.jneb.2005.11.021] [Medline: 16595275]
26. Wiedemann AU, Schüz B, Sniehotta F, Scholz U, Schwarzer R. Disentangling the relation between intentions, planning, and behaviour: a moderated mediation analysis. Psychol Health 2009 Jan;24(1):67-79. [doi: 10.1080/08870440801958214] [Medline: 20186640]
27. Varnhagen CK, Gushta M, Daniels J, Peters TC, Parmar N, Law D, et al. How informed is online informed consent? Ethics Behav 2005;15(1):37-48. [doi: 10.1207/s15327019eb1501_3] [Medline: 16127857]
28. Prochaska JO. Systems of Psychotherapy: A Transtheoretical Analysis. Homewood, IL: Dorsey Press; 1979.
29. Eat well. UK Food Standards Agency. Saturated Fat URL: http://www.eatwell.gov.uk/healthydiet/fss/fats/satfat/ [accessed 2010-02-01] [WebCite Cache ID 5nDvA2JkB]
30. Adriaanse MA, de Ridder DT, de Wit JB. Finding the critical cue: implementation intentions to change one's diet work best when tailored to personally relevant reasons for unhealthy eating. Pers Soc Psychol Bull 2009 Jan;35(1):60-71. [doi: 10.1177/0146167208325612] [Medline: 19106078]
31. Rossi SR, Greene GW, Rossi JS, Plummer BA, Benisovich SV, Keller S, et al. Validation of decisional balance and situational temptations measures for dietary fat reduction in a large school-based population of adolescents. Eat Behav 2001;2(1):1-18. [Medline: 15001046]
32. Velicer WF, Diclemente CC, Rossi JS, Prochaska JO. Relapse situations and self-efficacy: an integrative model. Addict Behav 1990;15(3):271-283. [Medline: 2378287]

XSL•FO
RenderX

33. Margetts BM, Cade JE, Osmond C. Comparison of a food frequency questionnaire with a diet record. Int J Epidemiol 1989 Dec;18(4):868-873. [Medline: 2621023]
34. Armitage CJ, Conner M. The theory of planned behaviour: assessment of predictive validity and 'perceived control'. Br J Soc Psychol 1999;38(1):35-54. [doi: 10.1348/014466699164022]
35. Thompson RL, Margetts BM. Comparison of a food frequency questionnaire with a 10-day weighed record in cigarette smokers. Int J Epidemiol 1993 Oct;22(5):824-833. [Medline: 8282461]
36. Renner B, Schwarzer R. International University Bremen & Freie Universität Berlin. Risk and Health Behaviors: Documentation of the Scales of the Research Project: "Risk Appraisal Consequences in Korea" (RACK). 2nd edition URL: http://www.gesundheitsrisiko.de/docs/RACKEnglish.pdf [accessed 2011-06-27] [WebCite Cache ID 5zkw6NCFj]
37. Sniehotta, FF, Scholz U, Schwarzer R. Bridging the intention-behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005;20(2):143-160. [doi: 10.1080/08870440512331317670]
38. Sniehotta FF, Scholz U, Schwarzer R. Action plans and coping plans for physical exercise: A longitudinal intervention study in cardiac rehabilitation. Br J Health Psychol 2006 Feb;11(Pt 1):23-37. [doi: 10.1348/135910705X43804] [Medline: 16480553]
39. Sniehotta FF, Schwarzer R, Scholz U, Schuz B. Action planning and coping planning for long-term lifestyle change: theory and assessment. Eur J Soc Psychol 2005;35(4):565-576. [doi: 10.1002/ejsp.258]
40. De Bourdeaudhuij I, Brug J. Tailoring dietary feedback to reduce fat intake: an intervention at the family level. Health Educ Res 2000 Aug;15(4):449-462 [FREE Full text] [Medline: 11066462]
41. Brug J, Glanz K, Van Assema P, Kok G, van Breukelen GJ. The impact of computer-tailored feedback and iterative feedback on fat, fruit, and vegetable intake. Health Educ Behav 1998 Aug;25(4):517-531. [Medline: 9690108]
42. Fair AK, Murray PG, Thomas A, Cobain MR. Using hypothetical data to assess the effect of numerical format and context on the perception of coronary heart disease risk. Am J Health Promot 2008;22(4):291-296. [doi: 10.4278/061030140R2.1] [Medline: 18421894]
43. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008 Aug;40(3):879-891. [Medline: 18697684]
44. World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2000 Dec 20;284(23):3043-3045. [Medline: 11122593]
45. van Sluijs EM, van Poppel MN, Twisk JW, van Mechelen W. Physical activity measurements affected participants' behavior in a randomized controlled trial. J Clin Epidemiol 2006 Apr;59(4):404-411. [doi: 10.1016/j.jclinepi.2005.08.016] [Medline: 16549263]
46. Nysveen H, Pedersen PE, Thorbjornsen H. Explaining intention to use mobile chat services: moderating effects of gender. Journal of Consumer MJ Consum Marketing 2005;22(5):247-256. [doi: 10.1108/07363760510611671]
47. Rafi MS. TESOL France. SMS Text Analysis: Language, Gender and Current Practices URL: http://www.tesol-france.org/Documents/Colloque07/SMS%20Text%20Analysis%20Language%20Gender%20and%20Current%20Practice%20_1_.pdf [accessed 2011-12-01] [WebCite Cache ID 63c4S3Fkl]
48. Minton HL, Schneider FW. Differential Psychology. Pacific Grove, CA: Brooks/Cole Publishing; 1980.
49. Rich L. The sociolinguistics of SMS: an analysis of SMS use by a random sample of Norwegians. Comput Support Coop Work 2008;31(4):335-349. [doi: 10.1007/1-84628-248-9_22]
50. Burger KS, Kern M, Coleman KJ. Characteristics of self-selected portion size in young adults. J Am Diet Assoc 2007 Apr;107(4):611-618. [doi: 10.1016/j.jada.2007.01.006] [Medline: 17383267]
51. Brunstrom JM, Rogers PJ, Pothos EM, Calitri R, Tapper K. Estimating everyday portion size using a 'method of constant stimuli': in a student sample, portion size is predicted by gender, dietary behaviour, and hunger, but not BMI. Appetite 2008 Sep;51(2):296-301. [doi: 10.1016/j.appet.2008.03.005] [Medline: 18467005]
52. Lippke S, Ziegelmann JP, Schwarzer R. Behavioral intentions and action plans promote physical exercise: a longitudinal study with orthopedic rehabilitation patients. J Sport Exerc Psychol 2004;26(3):470-483.
53. Hill C, Abraham C, Wright DB. Can theory-based messages in combination with cognitive prompts promote exercise in classroom settings? Soc Sci Med 2007 Sep;65(5):1049-1058. [doi: 10.1016/j.socscimed.2007.04.024] [Medline: 17544558]
54. Wiedemann AU, Lippke S, Reuter T, Ziegelmann JP, Schwarzer R. How planning facilitates behavior change: additive and interactive effects of a randomized controlled trial. Eur J Soc Psychol 2011;41(1):42-51. [doi: 10.1002/ejsp.724]
55. Garriguet D. Under-reporting of energy intake in the Canadian Community Health Survey. Health Rep 2008 Dec;19(4):37-45 [FREE Full text] [Medline: 19226926]
56. Cade JE, Burley VJ, Warm DL, Thompson RL, Margetts BM. Food-frequency questionnaires: a review of their design, validation and utilisation. Nutr Res Rev 2004 Jun;17(1):5-22. [doi: 10.1079/NRR200370] [Medline: 19079912]
57. Faith MS, Wong FY, Allison DB. Demand characteristics of the research setting can influence indexes of negative affect-induced eating in obese individuals. Obes Res 1998 Mar;6(2):134-136. [Medline: 9545020]
58. Aarts H, Dijkserhuis A, Midden C. To plan or not to plan? Goal achievement or interrupting the performance of a mundane behaviors. Eur J Soc Psychol 1999;29(8):971-979. [doi: 10.1002/(SICI)1099-0992(199912)29:8<971::AID-EJSP963>3.0.CO;2-A</971::AID-EJSP963>]

XSL•FO
RenderX

59. Webb TL, Sheeran P. How do implementation intentions promote goal attainment? A test of component processes. J Exp Soc Psychol 2007;43(2):295-302. [doi: doi:10.1016/j.jesp.2006.02.001]
60. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]

Abbreviations

PT: planning tool (condition)
PTT: planning tool and text messages (condition)

Edited by G Eysenbach; submitted 21.02.11; peer-reviewed by S Lippke; comments to author 14.03.11; revised version received 08.09.11; accepted 01.11.11; published 20.12.11.

Please cite as:
Soureti A, Murray P, Cobain M, Chinapaw M, van Mechelen W, Hurling R
Exploratory Study of Web-Based Planning and Mobile Text Reminders in an Overweight Population
J Med Internet Res 2011;13(4):e118
URL: http://www.jmir.org/2011/4/e118/
doi:10.2196/jmir.1773
PMID:22182483

©Anastasia Soureti, Peter Murray, Mark Cobain, Mai Chinapaw, Willem van Mechelen, Robert Hurling. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Tweets, Apps, and Pods: Results of the 6-Month Mobile Pounds Off Digitally (Mobile POD) Randomized Weight-Loss Intervention Among Adults

Gabrielle Turner-McGrievy[1], PhD, MS, RD; Deborah Tate[2], PhD

[1]Department of Health Promotion, Education, and Behavior, University of South Carolina, Columbia, SC, United States

[2]Department of Nutrition, Department of Health Behavior and Health Education, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Gabrielle Turner-McGrievy, PhD, MS, RD
Department of Health Promotion, Education, and Behavior
University of South Carolina
800 Sumter Street
Room 216
Columbia, SC, 29208
United States
Phone: 1 803 777 3932
Fax: 1 803 777 6290
Email: brie@sc.edu

Abstract

Background: Previous interventions have shown promising results using theory-based podcasts to deliver a behavioral weight-loss intervention.

Objective: The objective of our study was to examine whether a combination of podcasting, mobile support communication, and mobile diet monitoring can assist people in weight loss.

Methods: In this 6-month, minimal contact intervention, overweight (n = 96, body mass index 32.6 kg/m^2) adults were recruited through television advertisements and email listservs and randomly assigned to Podcast-only or Podcast+Mobile groups. Both groups received 2 podcasts per week for 3 months and 2 minipodcasts per week for months 3–6. In addition to the podcasts, the Podcast+Mobile group was also instructed to use a diet and physical activity monitoring application (app) on their mobile device and to interact with study counselors and other participants on Twitter.

Results: Weight loss did not differ by group at 6 months: mean –2.7% (SD 5.6%) Podcast+Mobile, n = 47; mean –2.7% (SD 5.1%) Podcast, n = 49; P = .98. Days/week of reported diet monitoring did not differ between Podcast+Mobile (mean 2.3, SD 1.9 days/week) and Podcast groups (mean 1.9, SD 1.7 days/week; P = .28) but method of monitoring did differ. Podcast+Mobile participants were 3.5 times more likely than the Podcast group to use an app to monitor diet (P = .01), whereas the majority of Podcast participants reported using the Web (14/41, 34%) or paper (12/41, 29%). There were more downloads per episode in the Podcast+Mobile group (1.4/person) than in the Podcast group (1.1/person; P < .001). The number of podcasts participants reported downloading over the 6-month period was significantly moderately correlated with weight loss in both the Podcast+Mobile (r = –.46, P = .001) and the Podcast (r = –.53, P < .001) groups. Podcast+Mobile participants felt more user control at 3 months (P = .02), but not at 6 months, and there was a trend (P = .06) toward greater elaboration among Podcast+Mobile participants. There were significant differences in reported source of social support between groups. More Podcast participants relied on friends (11/40, 28% vs 4/40, 10%; P = .045) whereas Podcast+Mobile participants relied on online sources (10/40, 25% vs 0/40; P = .001).

Conclusions: Results confirm and extend previous findings showing a minimally intensive weight-loss intervention can be delivered via podcast, but prompting and mobile communication via Twitter and monitoring app without feedback did not enhance weight loss.

Trial Registration: Clinicaltrials.gov NCT01139255; http://clinicaltrials.gov/ct2/show/NCT01139255 (Archived by WebCite at http://www.webcitation.org/625OjhiDy)

XSL•FO
RenderX

(J Med Internet Res 2011;13(4):e120) doi:10.2196/jmir.1841

KEYWORDS

Weight loss; obesity; podcasts as topic; social support

Introduction

The latest figures reveal that 68% of US adults are overweight or obese (body mass index [BMI] >25 kg/m^2), with the prevalence of obesity among adult women at 35.5% and among adult men at 32.2% [1]. Overweight and obesity is associated with several chronic diseases, including type 2 diabetes, hypertension, cardiovascular disease, arthritis, hyperlipidemia, and asthma [2,3]. Even modest weight loss—around a 5% decrease in body weight—has been shown to have significant impacts in the decrease of chronic disease risk [4,5].

Behavioral interventions that target improvements in diet and physical activity are an effective way to help people lose weight and decrease chronic disease risk factors [6]. Behavioral weight-loss research programs typically involve weekly or twice-monthly, face-to-face behavioral sessions involving a group of 10–20 members and a team of weight-loss research staff [7]. Although this can be an effective way to help people lose weight, it is time and resource intensive and not easy to disseminate. Additionally, many people feel that participation in face-to-face weight-loss interventions is time consuming and often inconvenient [8]. Moving away from a face-to-face setting to a mobile delivery method is a promising strategy in the delivery of a behavioral weight-loss intervention.

Mobile technologies, such as Internet-capable mobile devices (eg, iPhone and BlackBerry), could prove to be a useful conduit for delivery of a weight-loss program. The use of mobile devices has been on the rise. In 2010, 40% of adults in the United States used a mobile phone to access the Internet or send an email or instant message, and rates of these activities increased over the previous year [9]. US adults are also accessing audio using portable devices (MP3 players), with 46% of US adults reporting they own an MP3 player [9] and 19% of Internet users reporting that they have downloaded a podcast (an audio file that can be listened to on a computer or mobile media player) [10]. Podcasts may also see a growth in use due to the popularity of cloud computing (publishing, hosting, and accessing data all online), which has made the ability to listen to and create podcasts easier [11].

Research has been emerging on the use of mobile technologies to help people achieve a healthy weight. Mobile devices have been used successfully to provide dietary guidance [12] and self-monitor weight and other health-related variables [13,14] and dietary intake [15]. Few studies, however, have used an entirely mobile device-based approach to deliver a behavioral weight-loss intervention. For example, studies may use text messaging sent to mobile devices (short message service [SMS]) for support in addition to face-to-face group sessions [16]. One study that used an entirely mobile-based approach was conducted by Patrick and colleagues [17] among overweight men and women. This behavioral weight-loss intervention was delivered entirely through SMS with a mixture of standard and targeted messages to participants, who on average had lost 3.16% of their body weight at 4 months [17]. Another technology-delivered intervention that contained a sizable mobile component targeted increasing physical activity among healthy, slightly overweight (BMI 26.3 kg/m^2) men and women. Participants in the intervention group received targeted messages on overcoming barriers via the Internet and reminders to be physically active delivered via SMS and email, and had access to a message board to discuss experiences with other study participants [18].

Our previous work demonstrated that a short-term behavioral weight-loss intervention could be successfully delivered via podcast [19]. In that 3-month trial, 78 participants were randomly assigned to receive a podcast designed based on social cognitive theory [20] (enhanced) or a popular weight-loss podcast (control) available online. Weight losses were significantly greater in the enhanced podcast group (mean –2.9, SD 3.5 kg) than in the controls (mean –0.3, SD 2.1 kg; $P < .001$ between groups). This study, however, was short-term and weight losses were modest. A podcast-only format also limited the ability to provide participants with easy ways to self-monitor diet and physical activity and to receive social support (that would normally be delivered in a face-to-face group setting). To our knowledge, no previous studies have employed a combination of podcasting, mobile support communication, and mobile diet monitoring to assist people in weight loss. Therefore, we explored the use of this enhanced mobile approach as a way to help people lose weight in the Mobile Pounds Off Digitally (Mobile POD) intervention.

Methods

Study Population and Measures

Overweight and obese men and women (BMI 25–45 kg/m^2, 18–60 years old) were recruited through television advertisements and email listservs in the Raleigh-Durham, North Carolina, USA metropolitan area for this 6-month randomized trial. Participants were excluded if they smoked, had an unstable medical status or uncontrolled thyroid condition, were unable to attend the 3 monitoring visits or increase walking as a form of exercise, had a psychiatric illness, were in treatment for alcohol or drug dependency, had an eating disorder, were currently participating in a weight-loss program, or were pregnant, breastfeeding, or planning on becoming pregnant within the next 6 months. Participants were also required to be able to complete the Physical Activity Readiness Questionnaire [21] and were excluded for a history of myocardial infarction or stroke, and had to obtain physician consent for participation if endorsing yes on other items (such as use of hypertensive medications or bone and joint issues). Participants were required to have access to a body weight scale for self-monitoring weight and had to own one of four types of Internet-capable mobile devices: iPhone, iPod Touch, BlackBerry, or an Android-based

XSL•FO
RenderX

phone. Participants were required to have access to the Internet and be comfortable using a computer. The University of North Carolina at Chapel Hill Institutional Review Board approved the study, and all the participants gave written informed consent. Participants received a US $20 incentive payment for completing all 3- and 6-month assessment activities.

Study research assistants screened participants. On meeting screening criteria, participants were invited to an orientation session where they learned more about the study, were shown how to complete online baseline questionnaires, and filled out the consent form. Participants were then given 2 weeks to complete questionnaire items (all completed online) assessing the following: demographics; dietary intake from 2 days of unannounced 24-hour dietary recalls (1 weekday and 1 weekend day) collected using the Automated Self-administered 24-hour Dietary Recall [22]; physical activity (Paffenbarger Physical Activity Questionnaire) [23]; self-efficacy (Weight Efficacy Life-Style Questionnaire [WEL]) [24]; and eating behaviors (Eating Behavior Inventory [EBI]) [25]. Participants were also asked (on a 1- to 7-point Likert scale) to rate how supported they felt in their weight-loss efforts at 6 months. These questionnaire items were transformed from paper instruments to online methods (but had not been previously validated for online use).

Participants were randomly assigned using a computerized random numbers generator (as conducted by study investigators) once they completed of all their baseline questionnaires. A face-to-face group visit was scheduled to obtain objective height and weight. Before participants were given their randomization assignment, their weight was measured in light street clothes using a calibrated Tanita BWB-800 digital scale (Tanita, Arlington Heights, IL, USA) accurate to 0.1 kg. Height was measured using a wall-mounted stadiometer measured (Perspective Enterprises, Portage, MI, USA) after participants had removed hats and shoes. Once all baseline measures were collected, participants were given an overview of which group they were randomly assigned to and were provided with more details about group assignment. Both conditions were active treatments and participants were not told which group was the intervention of interest or enhanced group. Neither study participants nor investigators were blind to treatment assignment.

Intervention and Control Conditions

Participants could be randomly assigned to one of two conditions: podcast-only (Podcast) or podcast plus enhanced mobile media intervention (Podcast+Mobile). Both groups received 2 podcasts per week for 3 months (approximately 15 minutes each) and 2 minipodcasts per week for months 3–6 (approximately 5 minutes each). Participants had access to a group-specific podcast site, where they could subscribe to the podcast using their mobile device or listen directly to the podcast on a computer. The content and design of the podcasts have been described elsewhere [19]. Briefly, podcasts were designed using constructs from social cognitive theory [20]. Podcasts were written and recorded prior to the start of the study. Podcasts delivered in the first 3 months contained a section on nutrition and physical activity information, an audio blog of a man or a woman trying to lose weight, a soap opera, and a goal-setting activity. Podcasts delivered in months 3–6 contained only the nutrition and exercise portion of the podcast and focused on overcoming barriers and problem-solving issues. The Podcast group—but not the Podcast+Mobile group—received a book with calorie and fat gram amounts of food to assist them in monitoring their dietary intake. In addition to the podcasts, the Podcast+Mobile group was also instructed to download a diet and physical activity monitoring application (app) (2010 version of FatSecret's Calorie Counter app, FatSecret.com, which released additional updates in January 2011) and a social networking site's (Twitter) app to their mobile device (both free for download). Participants created a user account on Twitter, were told to log on (through either their mobile device or their computer) to Twitter at least once daily to read messages posted from the study coordinator, and were encouraged themselves to post at least daily to Twitter. Participants could choose any user name they wanted (to protect their identity) and were instructed on how to make their Twitter account private (if they chose to do so). During months 0–3, Podcast+Mobile participants were divided into 4 groups to create Twitter cohorts of 11–12 people. They were sent a list of everyone's Twitter user names within their cohort, were instructed to follow everyone in their cohort, and were reminded to send follow requests to participants and to accept requests until everyone in each cohort was following one another. During months 3–6, Podcast+Mobile participants were asked to follow everyone in the study, and similar procedures were used to allow everyone within the Podcast+Mobile group to follow one another. The study coordinator sent out 2 messages per day to the group, which reinforced messages from the podcasts, posed questions to the group to facilitate discussion, and encouraged participants to share tips and recipes with one another that would assist in weight loss. Such messages were prompts to attend to weight-loss behavior, and encouraged communication but were not individualized. The study coordinator did not participate in discussions initiated by participants. All participants received information on safe exercise practices.

Assessment Periods

Change in body weight was the main outcome of the study, and body weight was collected at baseline, 3 months, and 6 months at the study site. In addition to the diet, physical activity, and psychosocial measures discussed above, other measures were collected at both 3 and 6 months including novelty, cognitive load [26], user control [27], elaboration (Elaboration Likelihood Model Questionnaire) [28], and process evaluation questions, all via online questionnaire. Participants were also sent a weekly online questionnaire link so they could report the number of podcasts they had listened to that week, their weight, number of days they monitored their diet and physical activity, and, for the Podcast+Mobile group, questionnaire items assessing use of Twitter. Participants who did not complete their weekly questionnaire were sent an email reminder and received a phone reminder after 2 weeks in a row of missed weekly questionnaire submissions. The number of Twitter messages per participant was also recorded over the course of the study, and an objective measure of number of downloads per podcast by treatment group was obtained from the podcast hosting site.

XSL•FO
RenderX

Statistical Analyses

Power calculations for the study used the 3-month weight loss from the previous podcasting study [19] as compared with the 3-month weight loss seen in a Web-based study that used an automated email feedback group (similar to our Podcast+Mobile group) [29]. This resulted in an effect size of r = .2934 and a Cohen d of 0.6138. Sample size per intervention arm for 2-sided tests of significance at alpha = .05 and power 1 – beta = 80% would be 43 per group (86 total required N). To account for attrition, we attempted to recruit 95–105 total participants.

We conducted all data collection and analyses using intention-to-treat by using imputation (baseline observation carried forward), with the exception of some variables that we collected only at 6 months (such as information processing variables), which we assessed using completers only. Between-subjects t tests were calculated for differences between continuous variables, and paired-samples t tests were used to examine differences within groups. Logistic regression models were used to assess demographic predictors of study discontinuation at 6 months. Demographic information that contained multiple categories (such as education, ethnicity, and marital status) was dichotomized and the chi-square test of independence was used to assess differences between groups at baseline. Analysis of variance was used to examine mean differences within 3 or more groupings, and repeated-measures analysis of variance was used to assess changes over time among the continuous variables. Pearson correlation was used to examine the relationship between number of podcasts downloaded and the number of Twitter posts with weight loss. All analyses were conducted using SPSS 16.0 for Windows software (IBM Corporation, Somers, NY, USA) with a P value of .05 used to indicate statistically significant differences.

Results

Participants were recruited from July 2010 to August 2010, and the study ran until February 2011. Figure 1 shows the flow of participants through recruitment, intervention, and follow-up. Of the 494 volunteers who inquired about the study, 359 (72.7%) were ineligible and 135 were invited to an orientation, of whom 96 enrolled in the study. Table 1 outlines baseline demographics. There were no significant differences in baseline demographics between the two groups. More people in the Podcast+Mobile group than in the Podcast group reported previously downloading a health-related podcast (P = .04) or installing a healthy diet-related app to their mobile device (P = .04). In a model examining demographic factors as predictors of noncompletion of the study at 6 months, there was a significant effect of age (P = .005) and a trend with ethnicity (P = .06), but not of gender, group assignment, or baseline BMI. Participants who did not complete the study at 6 months (n = 10) were younger (mean 31.0, SD 11.2 vs 44.3, SD 10.3 years) and were 4.7 times (95% confidence interval, 0.96–23.24) less likely to be white (4/10, 40% black, 2/10, 20% Asian, and 4/10, 40% white).

XSL•FO
RenderX

Table 1. Baseline demographic data for Podcast-only and enhanced Podcast+Mobile group participants

	Podcast group (n = 49)	Podcast+Mobile group (n = 47)
Age (years), mean (SD)	43.2 (11.7)	42.6 (10.7)
Sex, n (%)		
Male	13 (27)	11 (23)
Female	36 (73)	36 (77)
Race/ethnicity, n (%)		
Black	10 (20)	9 (19)
White	38 (78)	35 (75)
Other	1 (2)	3 (6)
Hispanic, n (%)		
Yes	0	0
No	49 (100)	47 (100)
Marital status, n (%)		
Not married	23 (47)	16 (34)
Married	26 (53)	31 (66)
Education, n (%)		
College or less	19 (39)	24 (51)
Graduate degree	30 (61)	23 (49)
Type of Internet-capable mobile device, n (%)		
iPhone	14 (29)	18 (38)
iPod Touch	13 (26)	13 (28)
BlackBerry	18 (37)	11 (23)
Android-based phone	4 (8)	5 (11)
Body mass index (kg/m^2), mean (SD)	32.2 (4.5)	32.9 (4.8)
Number of years participant has owned Internet-capable mobile device, mean (SD)	1.6 (1.2)	1.3 (0.8)
Number of participants who were members of Twitter at baseline, n (%)	17 (35)	16 (34)
Number of participants who had previously downloaded a health-related podcast, n (%)	14 (29)[a]	23 (49)
Number of participants who had previously downloaded an application to their mobile device to help them eat better, n (%)	18 (37)[a]	27 (57)

[a] χ^2 test of independence $P = .04$.

Figure 2 details the weight loss by group over the 6-month study. The Podcast+Mobile group lost a mean of –2.4 (SD 3.4) kg at 3 months (vs –2.3, SD 3.3 kg in the Podcast group) and an additional –0.2 (SD 3.0) kg from months 3 to 6 (vs –0.3, SD 1.8 kg in the Podcast group; $P = .88$ for time-by-group interaction). Table 2 outlines group differences in percentage weight loss, diet, physical activity, self-efficacy, knowledge, and eating behaviors. The group-by-time interaction was not significant for any of the variables. The percentage weight loss at 3 or 6 months did not differ between the groups. There were no significant differences between groups in energy expenditure or intake at 3 or 6 months. Groups did not differ in changes in fat intake, self-efficacy (WEL score), or weight-related eating behaviors (EBI score) at 3 or 6 months.

XSL•FO
RenderX

Table 2. Changes in weight, physical activity, dietary intake, self-efficacy, knowledge, and eating behaviors at 3 and 6 months[a]

	Podcast group (n = 49)	Podcast+Mobile group (n = 47)	Significance (*P* value)
Weight change (%)			
0–3 months	–2.6 (3.8)	–2.6 (3.5)	.97
0–6 months	–2.7 (5.1)	–2.7 (5.6)	.98
Intentional physical activity change (caloric expenditure, in kcal)			
0–3 months	82.7 (153.2)	94.5 (130.2)	.68
0–6 months	96.7 (185.5)	86.8 (182.1)	.79
Change in energy intake (kcal)			
0–3 months	–146.3 (506.3)	–341.1 (612.1)	.09
0–6 months	–242.5 (558.8)	–288.8 (553.0)	.69
Change in fat intake (g)			
0–3 months	–13.6 (23.8)	–15.2 (31.0)	.78
0–6 months	–14.5 (32.0)	–15.0 (26.4)	.92
Change in weight-loss self-efficacy (WEL[b] score)			
0–3 months	12.5 (24.4)	12.5 (29.0)	.99
0–6 months	20.1 (26.0)	17.6 (25.3)	.64
Change in weight-loss knowledge score			
0–3 months	1.2 (1.8)	0.74 (1.9)	.24
0–6 months	1.1 (1.8)	0.66 (1.7)	.17
Change in eating behaviors (EBI[c] score)			
0–3 months	8.6 (12.6)	11.7 (11.8)	.21
0–6 months	9.8 (11.3)	12.4 (11.4)	.27

[a] All data are mean (SD).

[b] Weight Efficacy Life-Style Questionnaire.

[c] Eating Behavior Inventory.

Table 3 outlines differences in information processing variables measured at 3 and 6 months, as well as the number of podcasts downloaded and days per week diet and physical activity were self-monitored. Podcast+Mobile participants reported more user control at 3 months but not at 6 months (P = .08). There was no significant difference in cognitive load but Podcast+Mobile participants reported that the intervention was more novel at both 3 and 6 months. There was no significant difference in elaboration at 3 months between groups but there was a trend (P = .06) at 6 months with Podcast+Mobile participants reporting more elaboration than Podcast participants. There was no difference between groups at 0–3 months or 3–6 months in mean number of reported podcasts downloaded. However, the objective data from the podcast hosting site showed a significant difference in the number of downloads by group with more downloads occurring in the Podcast+Mobile group than in the Podcast group during both 0–3 months (P < .001) and 3–6 months (P < .001). The number of podcasts participants reported downloading over the 6-month period was significantly moderately correlated with weight loss in both the Podcast+Mobile (r = –.46, P = .001) and the Podcast (r = –.53, P < .001) groups. There was no difference between groups in the mean days per week that dietary intake or physical activity were self-reported by participants. The method of self-monitoring, however, differed by group. Podcast+Mobile participants were 3.5 times more likely than the Podcast group to use an app to monitor diet over the course of the study (95% confidence interval, 1.29–8.84; P = .01), whereas the majority of Podcast participants reported using Web (14/41, 34%) or paper (12/41, 29%) methods. Collapsing the data across groups, mean days per week of dietary self-monitoring over the 6-month study differed by method used to record intake: mean 2.9 (SD 1.9) days/week mobile app, n = 37; 2.3 (SD) 1.9 days/week website, n = 24; and 1.6 (SD 1.3) days/week paper journal, n = 17; n = 3 reported using nothing for monitoring and n = 15 did not report a method; $F_{2,7}$ = 3.41, P = .04). There was no difference in number of days of diet monitoring or weight loss by type of mobile device. Tukey honestly significant difference (HSD) post hoc analysis shows a significant difference in mean days per week of diet self-monitoring between paper journal methods and using a mobile app to record diet (P = .03).

XSL•FO
RenderX

Table 3. Information processing variables at 3 and 6 months and podcast downloads and self-monitoring outcomes by group[a]

	Podcast group (n = 49)	Podcast+Mobile group (n = 47)	Significance (*P* value)
User control [b]			
0–3 months	13.2 (5.0)	15.4 (3.5)	.02
3–6 months	13.9 (3.9)	15.4 (4.0)	.08
Cognitive load [b]			
0–3 months	11.7 (2.2)	11.5 (1.8)	.72
3–6 months	11.2 (2.6)	11.3 (2.4)	.79
Novelty [b]			
0–3 months	8.9 (3.1)	10.9 (3.0)	.01
3–6 months	9.1 (3.5)	11.1 (2.9)	.01
Elaboration [b]			
0–3 months	47.9 (10.0)	51.3 (8.4)	.1
3–6 months	45.7 (13.0)	50.3 (8.7)	.06
Mean number of podcasts (out of 24) participants reported they downloaded			
0–3 months	14.5 (7.6)	16.4 (7.2)	.20
3–6 months	8.2 (8.6)	9.0 (9.1)	.67
Number of downloads per podcast episode per person from podcasting host site			
0–3 months	1.51 (0.65)	2.00 (0.52)	<.001
3–6 months	0.66 (0.15)	0.87 (0.20)	<.001
Number of total podcast downloads from podcasting host site			
0–3 months	74.2 (31.8)	94.1 (24.6)	<.001
3–6 months	32.5 (7.2)	40.7 (9.5)	<.001
Mean days/week reported recording dietary intake			
0–3 months	2.4 (2.0)	2.9 (2.1)	.26
3–6 months	1.3 (1.7)	1.7 (2.0)	.39
Mean days/week reported recording physical activity			
0–3 months	2.6 (2.0)	2.4 (1.8)	.63
3–6 months	1.6 (2.1)	1.5 (1.9)	.81

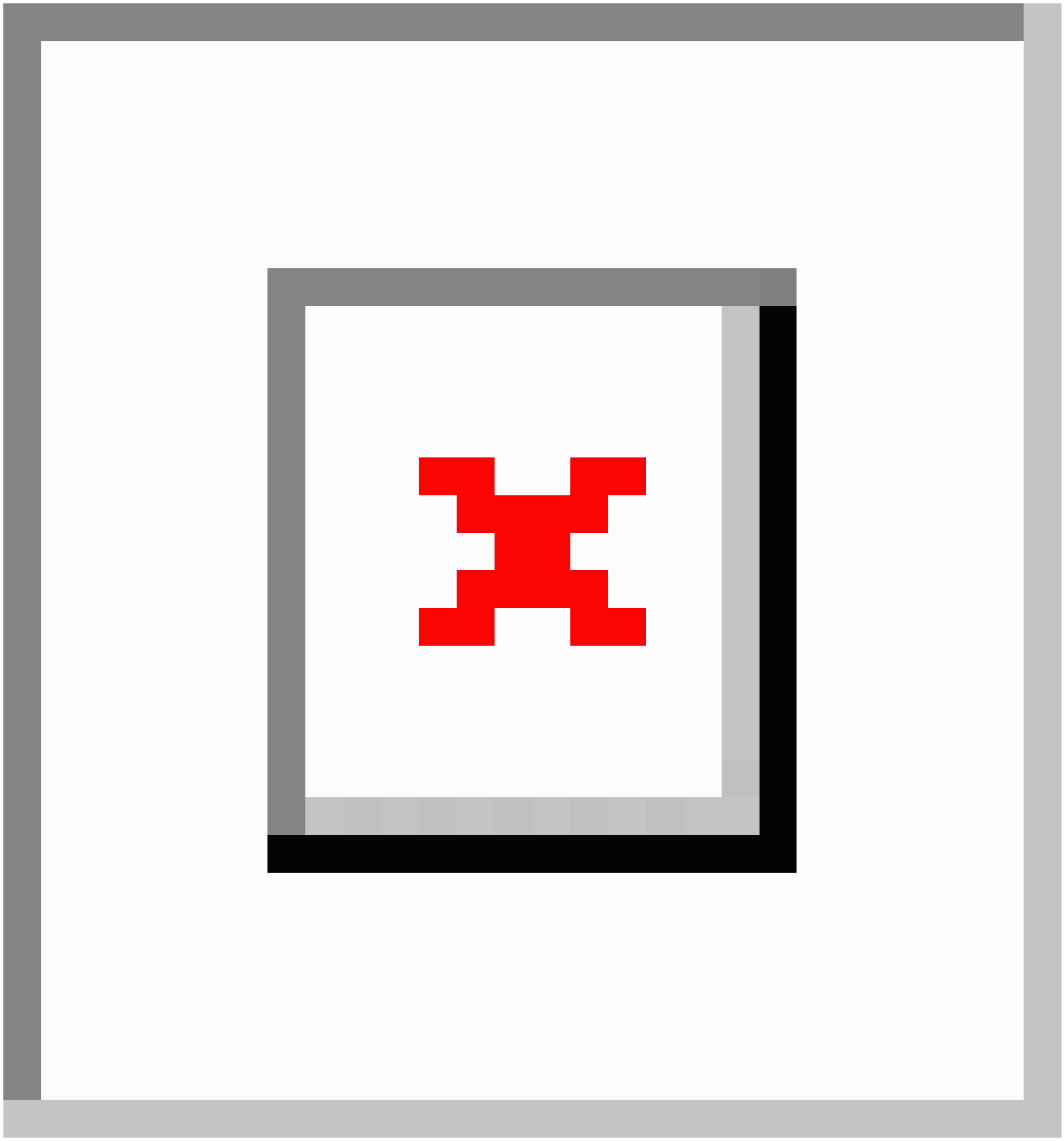
[a] All data are mean (SD).

[b] At 3 months, n = 43 for Podcast and n = 41 for Podcast+Mobile, and at 6 months n = 40 for both groups for all measures except elaboration at 6 months (n = 43 for Podcast and n = 40 for Podcast+Mobile).

Although there was no difference in how supported (rated 0–7) participants reported feeling at 6 months (Podcast+Mobile mean 5.0, SD 1.5 vs 4.8, SD 1.7; $P = .67$), there was a significant difference in the main form of social support participants reported during the 6-month trial. More of the Podcast participants reported mainly relying on friends for social support (11/40, 28% Podcast vs 4/40, 10% Podcast+Mobile; $P = .045$) and more Podcast+Mobile participants reported their main form of support came from online sources, such as Twitter, Facebook, or blogs (10/40, 25% Podcast+Mobile vs 0% Podcast; $P = .001$). Within the Podcast+Mobile group, 94% (n = 44) posted at least once to Twitter, with 64% (n = 30) posting at least weekly during the first 3 months and 28% (n = 13) posting weekly or more during months 3–6. Podcast+Mobile participants made a mean of 2.1 (SD 3.1) posts to Twitter per week, with significantly more posts being made in the first 3 months (2.8, SD 3.6 posts/week) than in months 3–6 (1.3, SD 3.0 posts/week; $P < .001$). On average, Podcast+Mobile group participants did not view Twitter as being useful to their weight-loss efforts (mean score of 3.4, SD 1.8, out of 7). Two technical issues occurred over the course of the study. FatSecret released an update to their app on January 3, 2011, which led to the app crashing for a few days before it was resolved by FatSecret. Also, on September 21, 2010, Twitter experienced a virus attack that was activated when users rolled their mouse over a blacked-out block of text [30]. It was quickly resolved by Twitter and no participant reported being affected by the virus (as it only affected Web users and not mobile app users).

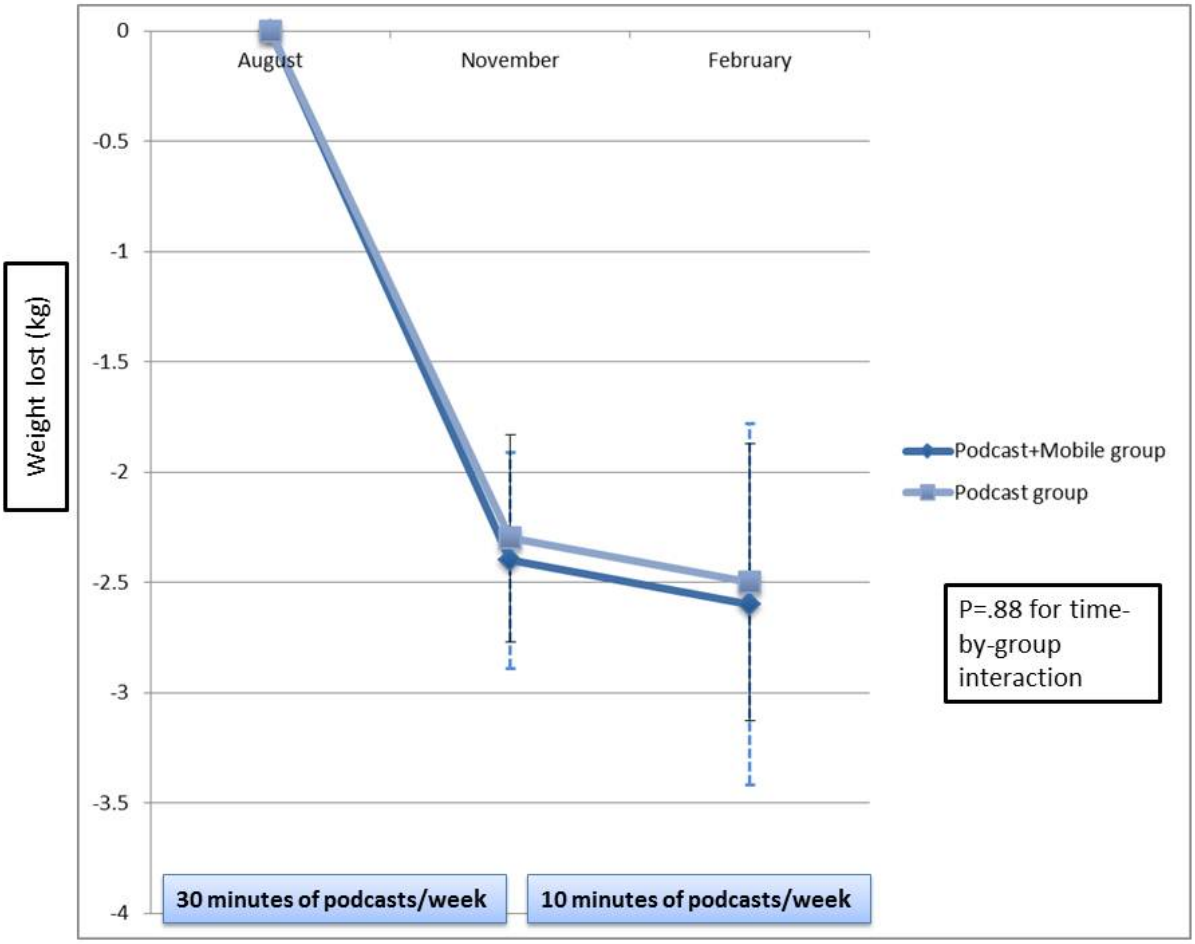
XSL•FO
RenderX

Figure 1. Eligibility, enrollment, random assignment, and assessment of study participants. BMI = body mass index.



XSL•FO
RenderX

Figure 2. Weight (mean, SE) lost by group over 6 months.



Discussion

This study explored the 6-month efficacy of a weight-loss intervention delivered by podcast and the additive benefit of mobile diet monitoring and communication delivered via a social networking site (Twitter). The results show that a very-low-intensity intervention that is delivered entirely by mobile technology can produce short-term, modest weight loss. The addition of prompts and support provided via Twitter, as well as mobile monitoring provided via a diet and physical activity app, did not enhance weight losses over what was seen by just using a podcast alone and encouraging participants to monitor with an approach of their choice.

Most participants in this study did not achieve a 5% weight loss, which is the level thought to be clinically meaningful; however, the time of year when the study was administered may have affected the outcomes. Participants completed the 3-month follow-up a week prior to the Thanksgiving holiday (United States). Months 3–6 occurred over the holiday season, including Thanksgiving, Christmas, Hanukkah, and New Year's. Many of the podcasts during this time discussed topics regarding holiday weight gain, such as altering holiday recipes, finding time to exercise when your schedule changes, and finding strategies to eat healthy at holiday parties. Observational studies have shown that the average weight gain over the holiday period in the United States ranges from 0.4 to 0.8 kg [31,32]. Holiday weight gain may be greater among those who are overweight or obese than among normal-weight people [31,32]. In addition, those who have lost weight are more vulnerable than normal-weight individuals to holiday weight gain and weight retention after the holiday [33]. Therefore, the fact that only minimal weight loss occurred during months 3–6 in the present study (approximately –0.25 kg or –0.55 lb) demonstrates that this intervention may have worked to prevent holiday weight gain, versus promoting additional weight loss during this vulnerable period. In addition, the podcasts during months 3–6 were shorter, only 5 minutes (10 minutes total/week), as compared with 15 minutes (30 minutes total/week) during months 0–3. It is possible that moving to a shorter intervention dose was insufficient to promote greater weight loss during this time. The combination of seasonality (holidays) and the lowering of the intervention dose at 3 months may have occurred too soon within the intervention, before participants had fully learned the behaviors needed for successful weight loss.

The intervention groups did not differ in changes in EBI score, WEL score, energy expenditure, or energy intake, demonstrating that the addition of social networking support and mobile diet and physical activity monitoring did not enhance these outcomes beyond receiving the theory-based podcast alone. Since there were no differences in days of self-reported diet monitoring by group and no differences in weight loss, these findings are not surprising. There were differences, however, in some of the information processing variables. Study interventionists

XSL•FO
RenderX

delivered 2 prompt-style messages to the Podcast+Mobile group via Twitter each day. We chose Twitter as a way to deliver real-time messages to participants from study interventionists (2 messages/day), which we hypothesized would be similar in effectiveness to delivering messages via SMS [17]. We also chose this social networking site to enhance elaboration, since messages posted by study staff reinforced messages delivered in the podcasts. Although elaboration did not differ between the groups at 3 months, there was a trend ($P = .06$) at 6 months, and it is possible that the addition of Twitter messages was beneficial in reinforcing weight loss-related messages and allowing for more effective information processing [34]. In addition to enhanced elaboration at 6 months, Podcast+Mobile participants reported greater user control at 3 months. Learning occurs in a different ways, so the more control learners have over their experience, the more variety of learning styles can be accommodated [35]. Therefore, providing additional learning channels (above audio alone) may have provided an additional feeling of user control and allowed participants to feel more motivated to learn and attend to messages [27]. Participants in both groups reported low levels of cognitive load (11 out of 14; with a higher score corresponding to less cognitive load) and therefore the addition of Twitter to the intervention did not increase the burden on working memory [36]. The addition of Twitter and mobile diet monitoring led to greater feelings of novelty among Podcast+Mobile participants than for the participants who received only a podcast. This enhanced novelty among Podcast+Mobile participants may have been due to the very low reported use of Twitter (34%) among participants at baseline. Despite changes in these information processing variables, weight losses did not differ between the groups. Greater elaboration and user control, however, may have led to greater podcast usage, as evidenced by more podcast downloads to the Podcast+Mobile group site. Greater usage of weight-loss study components has been shown to lead to improved weight losses [37]. Our prior podcasting study showed greater elaboration and user control among the enhanced theory-based podcast at 3 months than among the control podcast [19]. The present study also demonstrated high elaboration and user control within the Podcast+Mobile group as compared with the Podcast group, but no differences in cognitive load, demonstrating that the additional components of Twitter and mobile diet monitoring did not increase cognitive burden and allowed for continued elaboration and user control.

It is possible that the addition of Twitter and mobile monitoring was a distraction from what was already a successful weight-loss intervention delivered by podcast only. Because weight losses were modest in our previous podcasting trial [19], we sought to add components that are common to face-to-face behavioral interventions (self-monitoring, group support, and contact with study counselors) and deliver them in a mobile manner. These components appeared to enhance user control (at 3 months) without increasing cognitive load. The self-monitoring app and Twitter, however, were poorly used by participants. This demonstrates that these additional components were not well integrated by participants. In addition, the Podcast+Mobile participants reported relying more on online sources of support than on friends and family. There may have been a negative impact of displacing support from real-life friends and family with online social networks.

Self-monitoring of dietary intake is an important component of behavioral weight-loss programs [38]. We hypothesized that the use of a mobile diet monitoring app would increase dietary self-monitoring. We saw no differences in self-monitoring days per week at either 3 or 6 months between the groups, with both groups reporting monitoring an average of approximately 2.5 days/week from 0 to 3 months and 1.5 days/week from months 3 to 6. Although days per week of diet monitoring did not differ between groups, method of monitoring did. Podcast+Mobile participants were instructed to use the FatSecret Calorie Counter app. We chose this app due to its availability on all 3 major mobile phone platforms. Only 60% (24/40) of participants in the Podcast+Mobile group, however, reported using a mobile app for diet monitoring. The Podcast group was given a book to monitor their calorie intake, but at 6 months, only 29% (12/41) of participants were using a paper recording method to self-monitor dietary intake. Podcast participants may have been at an advantage with regard to diet monitoring, since they were able to choose which method they preferred to use and, if they chose to use an app (13/41, 32% of participants), they were free to choose which one would be best for their device. In a recent study examining differences in dietary intake between participants randomly assigned to monitor their diet via a handheld electronic device or paper journal, no differences were seen between the groups in weight loss, energy intake, or percentage of energy (kcal) from fat [39]. This demonstrates that adherence to monitoring is what is important for weight loss [40], regardless of method. With both groups collapsed, we did see a significant difference in number of days per week participants reported self-monitoring diet, with participants using a mobile device recording twice as many days as those using a paper method. This finding warrants further exploration in future studies and points to the possibility of recommending self-monitoring methods that are tailored to users' needs, mobile devices, and comfort level with technology.

The study content was delivered mainly through podcast messages for both groups. The number of podcasts participants reported downloading from months 0 to 3 and months 3 to 6 did not differ between the groups, but the objective number of downloads from each groups' podcast site did differ significantly. There were significantly more downloads per person in the Podcast+Mobile group than in the Podcast group. This may not have corresponded to more podcasts listened to, since participants may have initiated a download of an episode but later returned to finish the episode (starting another compete download). It is possible that Podcast+Mobile participants may have listened to podcast episodes multiple times as well or shared the podcast links with friends. Since Twitter messages prompted Podcast+Mobile participants to listen to the latest podcast episode, this method may have been an additional way for participants to remember to access the podcasts or be reminded to go back and listen to an episode as a refresher.

Social support has been shown to be a possible key component in behavioral weight-loss programs [41]. The present study used Twitter as a method to deliver prompts from the program and allow participants to support each other during their weight-loss

XSL•FO
RenderX

efforts. The study coordinator posted 2 messages to participants per day that could be easily automated, as the messages were not in response to participants. An additional advantage of Twitter is that it allowed for real-time support, such as asking fellow participants about healthy dining options once a participant arrives in a restaurant. However, the number of Twitter posts averaged 2 per week, and declined over time, so participants were not communicating with other group members frequently. Podcast+Mobile participants were initially assigned to a small cohort of 10–11 people. This was to allow for effective communication and to prevent participants from being overwhelmed with too many posts by other members. Participation was sporadic, however, so during the 3- to 6-month time period, Podcast+Mobile participants were asked to follow everyone within the Podcast+Mobile group to help facilitate more active discussion. It is possible there was not enough interaction to provide adequate social support, or this study may be similar to other studies that found no benefit of enhancing social support among weight-loss participants [42]. The podcasts, which were the same for both groups, encouraged establishing a good support system for weight loss. Both groups felt equally supported in their weight-loss efforts, but source of support differed. Podcast participants mainly turned to real-life friends as sources of support, whereas Podcast+Mobile participants relied on Internet-based sources, namely Twitter.

Several studies have shown that targeting dietary self-monitoring, providing social support, and emphasizing both dietary changes and physical activity are key components to successful face-to-face behavioral weight-loss programs [43]. We sought to improve on our previous trial [19] by enhancing these components through advanced mobile technology means. Both groups, however, found ways to self-monitor diet and obtain social support, meaning that structural aspects of the groups differed in type but not amount, which was reflected in the weight-loss outcomes. Several of the components of Internet-based weight-loss interventions that have been shown to be effective were part of the Podcast+Mobile intervention, including self-monitoring, use of established behavioral strategies, and social support [44]. Other effective aspects were not included, such as study counselor feedback and tailoring of messages and information [44]. Future studies examining ways to enhance weight loss of participants receiving a mobile, podcast-delivered weight-loss intervention may wish to find ways to provide more individualized feedback (which could be human or automated) on dietary self-monitoring and intake, physical activity progress, and weight loss. Additionally, ways to tailor the intervention components to participant characteristics, such as body weight, dietary intake, or physical activity, may enhance weight losses over nontailored approaches [44]. Designers of future interventions may wish to find ways to offer a multimodal intervention to appeal to diverse users.

There were several strengths to this study. This intervention used a randomized design and intention-to-treat analysis. The study also had strong retention rates, used an objective measure for weight, and included 2 unannounced days of dietary intake collected using a multiple-pass method. This intervention is an approach that could be easily disseminated, as there was very minimal contact with participants, everything was delivered remotely, and it was low cost. The study also had some limitations. The study population was mostly white and female. The study was short-term (6 months) and so it was able to examine initial weight loss but not weight-loss maintenance. The groups also differed in prior experience with apps and podcasts, showing that those in the Podcast+Mobile group may have been more technically savvy or more familiar with the technology used in the study at baseline. This did not seem to affect the results, and equal numbers of participants from each group were members of Twitter at baseline. Participants who did not complete the study were significantly younger and less likely than completers to be white, and therefore some aspect of the study may not have been well suited to this demographic group.

In summary, the Mobile POD, 6-month, minimal-contact intervention was effective at helping participants achieve a mean weight loss of 2.7% of their body weight, and perhaps was useful in preventing holiday weight gain. Both groups reported similar levels of social support and days of dietary monitoring, demonstrating that providing these components was not necessary, and that the podcasts' emphasis on directing participants to find these components was effective. Future studies may wish to find ways to combine podcasts with tailored feedback for participants to enhance compliance with dietary and physical activity recommendations and to improve weight loss.

Acknowledgments

The authors would like to thank the UNC Lineberger Comprehensive Cancer Center Population Sciences Award and the UNC Interdisciplinary Obesity Center for providing funding for this research.

Conflicts of Interest

None declared

References

1. Flegal KM, Carroll MD, Ogden CL, Curtin LR. Prevalence and trends in obesity among US adults, 1999-2008. JAMA 2010 Jan 20;303(3):235-241. [doi: 10.1001/jama.2009.2014] [Medline: 20071471]
2. Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. JAMA 2003 Jan 1;289(1):76-79 [FREE Full text] [Medline: 12503980]

XSL•FO
RenderX

3. Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The disease burden associated with overweight and obesity. JAMA 1999 Oct 27;282(16):1523-1529 [FREE Full text] [Medline: 10546691]
4. Blackburn G. Effect of degree of weight loss on health benefits. Obes Res 1995 Sep;3 Suppl 2:211s-216s. [Medline: 8581779]
5. Pasanisi F, Contaldo F, de Simone G, Mancini M. Benefits of sustained moderate weight loss in obesity. Nutr Metab Cardiovasc Dis 2001 Dec;11(6):401-406. [Medline: 12055705]
6. Dombrowski SU, Avenell A, Sniehott FF. Behavioural interventions for obese adults with additional risk factors for morbidity: systematic review of effects on behaviour, weight and disease risk factors. Obes Facts 2010 Dec;3(6):377-396. [doi: 10.1159/000323076] [Medline: 21196792]
7. Smith CF, Wing RR. New directions in behavioral weight-loss programs. Diabetes Spectr 2000;13(3):142-148.
8. Sherwood NE, Morton N, Jeffery RW, French SA, Neumark-Sztainer D, Falkner NH. Consumer preferences in format and type of community-based weight control programs. Am J Health Promot 1998;13(1):12-18. [Medline: 10186930]
9. Smith A. Mobile access 2010. Washington, DC: Pew Internet & American Life Project; 2010 Jul 07. URL: http://pewinternet.org/~/media//Files/Reports/2010/PIP_Mobile_Access_2010.pdf [accessed 2011-12-02] [WebCite Cache ID 63djQMKar]
10. Madden M, Jones S. Podcast downloading: 19% of Internet users have downloaded a podcast. Washington, DC: Pew Internet & American Life Project; 2008 Aug 28. URL: http://www.pewinternet.org/~/media//Files/Reports/2008/PIP_Podcast_2008_Memo.pdf.pdf [accessed 2011-12-02] [WebCite Cache ID 63djVBvh2]
11. Madden M. Podcasting into the cloud. Washington, DC: Pew Internet & American Life Project; 2008 Sep 12. URL: http://www.pewinternet.org/Commentary/2008/September/Podcasting-into-the-Cloud.aspx [accessed 2011-12-02] [WebCite Cache ID 63djZmkrA]
12. Rossi MC, Perozzi C, Consorti C, Almonti T, Foglini P, Giostra N, et al. An interactive diary for diet management (DAI): a new telemedicine system able to promote body weight reduction, nutritional education, and consumption of fresh local produce. Diabetes Technol Ther 2010 Aug;12(8):641-647. [doi: 10.1089/dia.2010.0025] [Medline: 20615106]
13. Lee W, Chae YM, Kim S, Ho SH, Choi I. Evaluation of a mobile phone-based diet game for weight control. J Telemed Telecare 2010 Jul;16(5):270-275. [doi: 10.1258/jtt.2010.090913] [Medline: 20558620]
14. Morak J, Schindler K, Goerzer E, Kastner P, Toplak H, Ludvik B, et al. A pilot study of mobile phone-based therapy for obese patients. J Telemed Telecare 2008;14(3):147-149. [doi: 10.1258/jtt.2008.003015] [Medline: 18430285]
15. Anhøj J, Møldrup C. Feasibility of collecting diary data from asthma patients through mobile phones and SMS (short message service): response rate analysis and focus group evaluation from a pilot study. J Med Internet Res 2004 Dec 2;6(4):e42 [FREE Full text] [doi: 10.2196/jmir.6.4.e42] [Medline: 15631966]
16. Lombard C, Deeks A, Jolley D, Ball K, Teede H. A low intensity, community based lifestyle programme to prevent weight gain in women with young children: cluster randomised controlled trial. BMJ 2010;341:c3215 [FREE Full text] [Medline: 20627974]
17. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]
18. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
19. Turner-McGrievy GM, Campbell MK, Tate DF, Truesdale KP, Bowling JM, Crosby L. Pounds Off Digitally study: a randomized podcasting weight-loss intervention. Am J Prev Med 2009 Oct;37(4):263-269. [doi: 10.1016/j.amepre.2009.06.010] [Medline: 19765496]
20. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
21. Cardinal BJ, Esters J, Cardinal MK. Evaluation of the revised physical activity readiness questionnaire in older adults. Med Sci Sports Exerc 1996 Apr;28(4):468-472. [Medline: 8778552]
22. Subar AF, Crafts J, Zimmerman TP, Wilson M, Mittl B, Islam NG, et al. Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. J Am Diet Assoc 2010 Jan;110(1):55-64. [doi: 10.1016/j.jada.2009.10.007] [Medline: 20102828]
23. Paffenbarger RS, Wing AL, Hyde RT, Jung DL. Physical activity and incidence of hypertension in college alumni. Am J Epidemiol 1983 Mar;117(3):245-257. [Medline: 6829553]
24. Clark MM, Abrams DB, Niaura RS, Eaton CA, Rossi JS. Self-efficacy in weight management. J Consult Clin Psychol 1991 Oct;59(5):739-744. [Medline: 1955608]
25. O'Neil PM, Rieder S. Utility and validity of the eating behavior inventory in clinical obesity research: a review of the literature. Obes Rev 2005 Aug;6(3):209-216. [doi: 10.1111/j.1467-789X.2005.00192.x] [Medline: 16045636]
26. Brunken R, Plass JL, Leutner D. Direct measurement of cognitive load in multimedia learning. Educ Psychol 2003;38(1):53-61. [doi: 10.1207/S15326985EP3801_7]
27. Eveland W, Dunwoody S. User control and structural isomorphism or disorientation and cognitive load? Learning from the Web versus print. Commun Res 2001;28(1):48-78. [doi: 10.1177/009365001028001002]

XSL•FO
RenderX

28. Heppner MJ, Humphrey CF, Hillenbrand-Gunn TL, DeBord KA. The differential effects of rape prevention programming on attitudes, behavior, and knowledge. J Couns Psychol 1995;42(4):508-518. [doi: 10.1037/0022-0167.42.4.508]
29. 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 [FREE Full text] [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
30. Ortutay B. The Huffington Post. 2010 Sep 21. Twitter HACKED: Twitter Virus Opens Popups, Causes Havoc URL: http://www.huffingtonpost.com/2010/09/21/twitter-hacked-twitter-vi_n_733338.html [accessed 2011-09-30] [WebCite Cache ID 625QGB9yK]
31. Hull HR, Radley D, Dinger MK, Fields DA. The effect of the Thanksgiving holiday on weight gain. Nutr J 2006;5:29 [FREE Full text] [doi: 10.1186/1475-2891-5-29] [Medline: 17118202]
32. Yanovski JA, Yanovski SZ, Sovik KN, Nguyen TT, O'Neil PM, Sebring NG. A prospective study of holiday weight gain. N Engl J Med 2000 Mar 23;342(12):861-867 [FREE Full text] [doi: 10.1056/NEJM200003233421206] [Medline: 10727591]
33. Phelan S, Wing RR, Raynor HA, Dibello J, Nedeau K, Peng W. Holiday weight management by successful weight losers and normal weight individuals. J Consult Clin Psychol 2008 Jun;76(3):442-448. [doi: 10.1037/0022-006X.76.3.442] [Medline: 18540737]
34. Cialdini RB, Petty RE, Cacioppo JT. Attitude and attitude change. Annu Rev Psychol 1981;32:357-404. [doi: 10.1146/annurev.ps.32.020181.002041]
35. Ariely D. Controlling the information flow: effects on consumers' decision making and preferences. J Consum Res 2000;27(2):233-248 [FREE Full text]
36. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. Educ Psychol 2003;38(1):63-71. [doi: 10.1207/S15326985EP3801_8]
37. Webber KH, Tate DF, Michael Bowling J. A randomized comparison of two motivationally enhanced Internet behavioral weight loss programs. Behav Res Ther 2008 Sep;46(9):1090-1095. [doi: 10.1016/j.brat.2008.06.008] [Medline: 18675402]
38. Burke LE, Warziski M, Starrett T, Choo J, Music E, Sereika S, et al. Self-monitoring dietary intake: current and future practices. J Ren Nutr 2005 Jul;15(3):281-290. [Medline: 16007557]
39. Acharya SD, Elci OU, Sereika SM, Styn MA, Burke LE. Using a personal digital assistant for self-monitoring influences diet quality in comparison to a standard paper record among overweight/obese adults. J Am Diet Assoc 2011 Apr;111(4):583-588. [doi: 10.1016/j.jada.2011.01.009] [Medline: 21443993]
40. Hollis JF, Gullion CM, Stevens VJ, Brantley PJ, Appel LJ, Ard JD, Weight Loss Maintenance Trial Research Group. Weight loss during the intensive intervention phase of the weight-loss maintenance trial. Am J Prev Med 2008 Aug;35(2):118-126. [doi: 10.1016/j.amepre.2008.04.013] [Medline: 18617080]
41. Verheijden MW, Bakx JC, van Weel C, Koelen MA, van Staveren WA. Role of social support in lifestyle-focused weight management interventions. Eur J Clin Nutr 2005 Aug;59 Suppl 1:S179-S186. [doi: 10.1038/sj.ejcn.1602194] [Medline: 16052189]
42. Kumanyika SK, Wadden TA, Shults J, Fassbender JE, Brown SD, Bowman MA, et al. Trial of family and friend support for weight loss in African American adults. Arch Intern Med 2009 Oct 26;169(19):1795-1804 [FREE Full text] [doi: 10.1001/archinternmed.2009.337] [Medline: 19858438]
43. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
44. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938. [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]

Abbreviations

app: application
BMI: body mass index
EBI: Eating Behavior Inventory
Mobile POD: Mobile Pounds Off Digitally
SMS: short message service
WEL: Weight Efficacy Life-Style Questionnaire

XSL•FO
RenderX

Edited by G Eysenbach; submitted 06.05.11; peer-reviewed by R Hurling; comments to author 26.09.11; revised version received 07.10.11; accepted 17.10.11; published 20.12.11.

Please cite as:
Turner-McGrievy G, Tate D
Tweets, Apps, and Pods: Results of the 6-Month Mobile Pounds Off Digitally (Mobile POD) Randomized Weight-Loss Intervention Among Adults
J Med Internet Res 2011;13(4):e120
URL: http://www.jmir.org/2011/4/e120/
doi:10.2196/jmir.1841
PMID:22186428

©Gabrielle Turner-McGrievy, Deborah Tate. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Web-Based Versus Face-To-Face Delivery of Education in Prescription of Falls-Prevention Exercise to Health Professionals: Randomized Trial

Stephen Maloney[1], BApplSc (Physio), MPH; Romi Haas[2], BApplSc (Physio), MPH; Jennifer L Keating[1], PhD; Elizabeth Molloy[3], PhD; Brian Jolly[4], PhD; Jane Sims[5], PhD; Prue Morgan[1], BApplSc (Physio), MApplSc; Terry Haines[1,2], PhD

[1]Department of Physiotherapy, Monash University, Frankston, Australia

[2]Southern Health, Melbourne, Australia

[3]Health Professions Education and Educational Research, Monash University, Melbourne, Australia

[4]Health Workforce Education and Assessment Research, Monash University, Melbourne, Australia

[5]Healthy Ageing Research Unit, Monash University, Melbourne, Australia

Corresponding Author:
Stephen Maloney, BApplSc (Physio), MPH
Department of Physiotherapy
Monash University
PO Box 527
Frankston, 3199
Australia
Phone: 61 3 99044240
Fax: 61 990 44812
Email: stephen.maloney@monash.edu

Abstract

Background: Exercise is an effective intervention for the prevention of falls; however, some forms of exercises have been shown to be more effective than others. There is a need to identify effective and efficient methods for training health professionals in exercise prescription for falls prevention.

Objective: The objective of our study was to compare two approaches for training clinicians in prescribing exercise to prevent falls.

Methods: This study was a head-to-head randomized trial design. Participants were physiotherapists, occupational therapists, nurses, and exercise physiologists working in Victoria, Australia. Participants randomly assigned to one group received face-to-face traditional education using a 1-day seminar format with additional video and written support material. The other participants received Web-based delivery of the equivalent educational material over a 4-week period with remote tutor facilitation. Outcomes were measured across levels 1 to 3 of Kirkpatrick's hierarchy of educational outcomes, including attendance, adherence, satisfaction, knowledge, and self-reported change in practice.

Results: Of the 166 participants initially recruited, there was gradual attrition from randomization to participation in the trial (n = 67 Web-based, n = 68 face-to-face), to completion of the educational content (n = 44 Web-based, n = 50 face-to-face), to completion of the posteducation examinations (n = 43 Web-based, n = 49 face-to-face). Participant satisfaction was not significantly different between the intervention groups: mean (SD) satisfaction with content and relevance of course material was 25.73 (5.14) in the Web-based and 26.11 (5.41) in the face-to-face group; linear regression $P = .75$; and mean (SD) satisfaction with course facilitation and support was 11.61 (2.00) in the Web-based and 12.08 (1.54) in the face-to-face group; linear regression $P = .25$. Knowledge test results were comparable between the Web-based and face-to-face groups: median (interquartile range [IQR]) for the Web-based group was 90.00 (70.89–90.67) and for the face-to-face group was 80.56 (70.67–90.00); rank sum $P = .07$. The median (IQR) scores for the exercise assignment were also comparable: Web-based, 78.6 (68.5–85.1), and face-to-face, 78.6 (70.8–86.9); rank sum $P = .61$. No significant difference was identified in Kirkpatrick's hierarchy domain *change in practice*: mean (SD) Web-based, 21.75 (4.40), and face-to-face, 21.88 (3.24); linear regression $P = .89$.

XSL•FO
RenderX

Conclusion: Web-based and face-to-face approaches to the delivery of education to clinicians on the subject of exercise prescription for falls prevention produced equivalent results in all of the outcome domains. Practical considerations should arguably drive choice of delivery method, which may favor Web-based provision for its ability to overcome access issues for health professionals in regional and remote settings.

Trial Registration: Australian New Zealand Clinical Trials Registry number: ACTRN12610000135011; http://www.anzctr.org.au/ACTRN12610000135011.aspx (Archived by WebCite at http://www.webcitation.org/63MicDjPV)

(J Med Internet Res 2011;13(4):e116) doi:10.2196/jmir.1680

KEYWORDS

Education; professional development; course design; distance education; students

Introduction

Continuing professional development (CPD) is an essential component of the educational continuum of health care professionals. Participation in continuing education is a hallmark of professional behavior [1,2] and a requirement for continued registration to practice for many professions. It is an imperative for continued advancement in the quality of health care and is a key educational activity for universities, professional organizations, and other educational institutions [2]. Effectively delivered CPD can narrow the gap between new evidence of best practice and current practice, enhancing practitioners' knowledge and behaviors and positively influencing service delivery and patient outcomes [3-5].

Despite the importance of CPD, its uptake is impeded by personal, professional, and environmental factors [6]. Previously reported barriers to participating in CPD have included time, access, and cost [7]. Clinicians have expressed reluctance to make time for CPD (due to the demands of busy caseloads, social life and family, and travel to education-delivery venues) and have noted constraints associated with the direct costs of CPD and indirect costs associated with lost earnings. These factors are amplified in regional and remote areas where workforce shortages can increase caseloads, and geographic distance renders CPD less accessible and more costly [8]. These constraining factors may have a negative effect on workforce satisfaction and make it difficult to retain staff in regional and remote areas. Clinicians in regional and remote areas of Australia have reported dissatisfaction with lack of professional development opportunities and the associated professional isolation, and these factors have been linked to staff recruitment and retention difficulties [9].

The traditional approach to CPD has been face-to-face education, often involving large live audiences (eg, seminars or conference presentations). This face-to-face approach may be difficult or costly to access for people who live a long distance from the education venue. Further to this, participants need to be able to allocate the requisite time in their schedule to enable participation. This mode of educational delivery can create a significant burden on educational institutions through administrative load and financial risk if the threshold for participant numbers required for financial viability is not reached. These are key factors that limit the implementation of training programs for health professionals in regional and remote areas.

Web-based CPD provides an alternative method of education with the potential to overcome many reported obstacles. Web-based delivery of educational content provides flexibility of access and promotes a learner-centered approach to learning, enabling interaction with learning materials at a time that suits the consumer [10]. Using the Web to deliver CPD requires the consumer to have Internet access, information technology literacy, and occasional technical support, as well as the evolution of teaching and learning resources that are tailored to suit the medium [11,12].

It is possible that multimedia or Web-based instruction may be inappropriate for teaching or monitoring practical skills [11-14]. Practical-skills competence requires a degree of self-evaluation that is often inadequate in the absence of appropriate feedback [15]. However, innovative design of feedback using interactive and collaborative Web-based technology might overcome the purported limitations [16].

Previous investigations of Web-based education have contrasted this with providing no education [17-27], rather than applying a head-to-head approach for comparing Web-based with face-to-face education. One report of an investigation comparing behavior change from Web-based versus face-to-face education found that Web-based CPD can produce changes in behavior in physicians working in lipid management, when practice change is measured using the number of requests for lipid management tests [28]. Gains in knowledge comparable with or superior to those obtained via live education delivery were also seen. It is plausible that Web-based delivery can also produce learning outcomes equal to face-to-face education when applied to the delivery of a complex hands-on clinical skill.

In this study we compared a Web-based approach to providing a CPD course versus a traditional "live" education-delivery approach in the context of education for the complex clinical skill of exercise prescription for falls prevention. Falls are a significant threat to the safety, health, and independence of older adults, accounting for more than half of the accidental deaths among older adults [29,30]. In Australia, the total cost of fall-related injury is expected to triple to A$1375 million per year by 2051 unless effective prevention or lower treatment costs occur [31]. Exercise is an effective intervention for the prevention of falls; however, some forms of exercises have been shown to be more effective than others, with systematic reviews and meta-analyses finding high-dose exercise, including challenging balance training, the most effective [32,33]. Several Web-based falls-prevention programs have been introduced

XSL•FO
RenderX

[34,35]. Some exercise programs, such as the Otago program, have been the focus of Australian nationwide public health strategies to address the issue. Programs to date have focused on the provision of one single program, and not on teaching the learner the broad range of exercise prescription skills that would allow exercises to be designed, tailored, and applied by the clinician to those at risk of falls. Exercise prescription is a skill set that requires knowledge of anatomy, biomechanics, psychology, and the practical skills to safely guide the patient to achieve targets of improved balance, stability, risk reduction, and confidence with ambulation. Teaching practical skills such as guiding a patient to master exercise for the prevention of falls provides an important and clinically relevant context for investigating the relative effectiveness of Web-based and face-to-face modalities.

Methods

Design

This study was a randomized trial with concealed allocation and blind outcome assessment comparing two educational interventions (Web-based compared with face-to-face education in falls-prevention intervention) that employed a mixed (qualitative and quantitative) evaluation framework. Ethics approval was obtained through both the Monash University Human Research Ethics Committee and the Southern Health Ethics Committee.

Participants

Participants were required to hold at least a bachelor's degree in any health science and reside in the state of Victoria, Australia. Participants were invited through a recruitment information package consisting of an electronic flyer, explanatory statement, consent form, and registration form. The recruitment package was distributed by email through managers of target professional disciplines via the Victorian department of health email distribution channels and by direct contact with private practices and community health centers in the geographic areas, two regional and one metropolitan, where face-to-face delivery was scheduled. Registrations were accepted for a period of 6 weeks from the time of the mailout, closing 1 week before the intervention started (May 2010). Applications for registration were screened for evidence that inclusion criteria were met (bachelor's degree in a health science and residing in Victoria, Australia).

Interventions

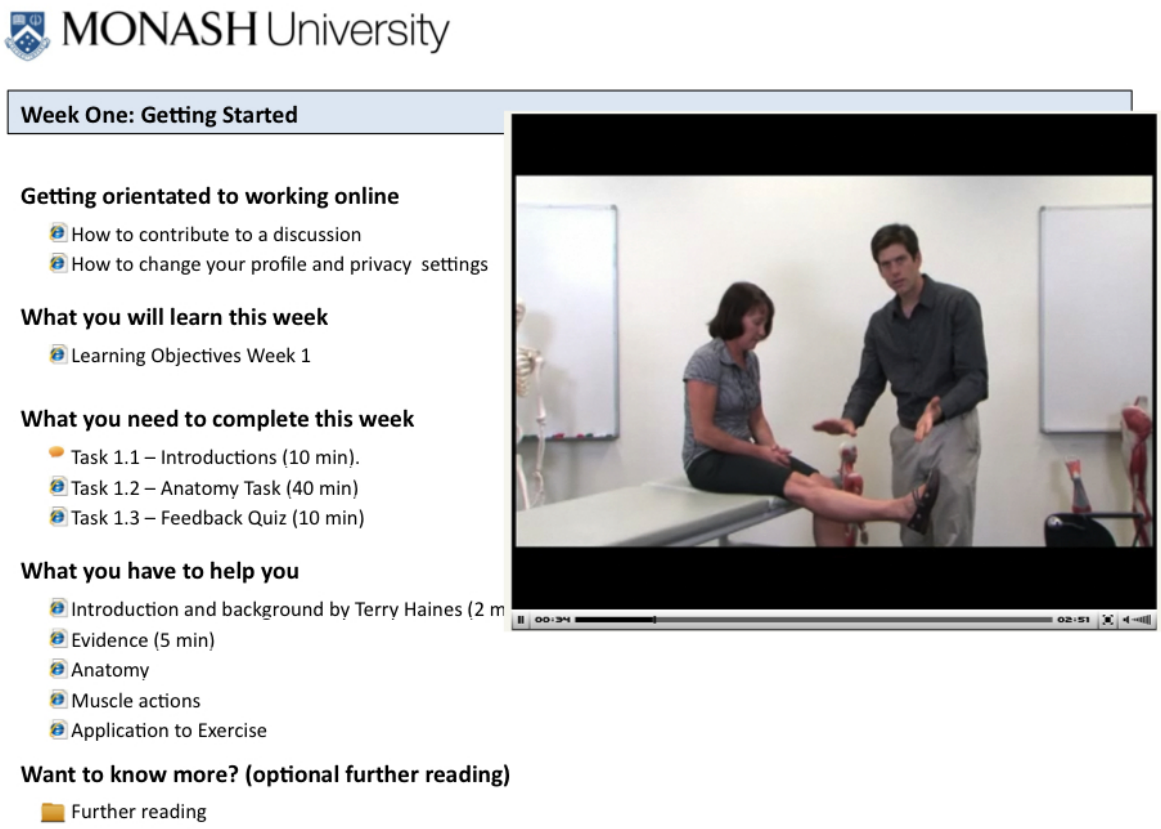
The content of the interventions was informed by three scoping activities: a review of the falls-prevention program literature to establish common elements in existing falls-education programs; phone interviews conducted with 24 clinicians who were leaders in falls prevention to establish current practice in specialized clinics; and phone interviews conducted with six target audience representatives from multiple professions to identify the knowledge gap between expert clinicians and the education target audience. The course content then underwent an external review by a falls-prevention researcher, a clinical leader, an academic leader, and a member of the falls-risk target audience. All learning objectives for both interventions were mapped and linked to relevant resources and tasks, and were matched in content and time requirements. The learning objectives included the physiological principles of exercise prescription for falls prevention, assessment procedures, exercise selection and delivery, and techniques to encourage program adherence and behavior change. To enhance consistency of delivery and adherence to the planned curriculum, the face-to-face leader, his or her delivery assistant, and the Web-based facilitator were all trained from a single DVD, comprising the same footage and learning resources used in the construction of the Web-based program.

One group received face-to-face education by attending a 1-day (7-hour) seminar, facilitated by a local expert in this field who had practiced clinically in the area, and had published and completed a PhD in the arena of falls prevention. To assist student revision, a support package was posted to participants before the seminar started that included a copy of the presentation slides, reference to further readings, and a DVD of the assessment procedures to be covered in the seminar. The seminar was held outside usual work hours with participants allocated to the nearest program location.

The other group took part in Web-based delivery of the equivalent educational material over a 4-week period (anticipated to require 7 hours total time commitment over 4 weeks) facilitated by a Web-based tutor who corresponded with participants through Web-based discussions and was available by phone if problems occurred. The Web-based course was constructed within the online learning system Moodle (Moodle.com, Perth, Australia), which uses open source code and is available in the public domain. Figure 1 shows a screenshot of the constructed course home page to illustrate the typical integration of activities and learning resources. Participants were posted a DVD comprising the multimedia used in the Web-based program as a troubleshooting solution if they encountered difficulties viewing the content online. Participants were allowed to progress through the program at their own pace, completing educational activities any time during the 4 weeks. Learning tasks ranged from self-directed reading and formative quizzes to interactive skills-practice sessions with feedback opportunities. For feedback, students uploaded digital footage of their skill mastery, which was viewed by the Web-based tutor. They were then guided through a reflective task by reading the tutor's comments of typical group performance in the task submissions, and they could view a tutor-selected exemplar of student performance to enable benchmarking of expectations of performance competency.

XSL•FO
RenderX

Figure 1. Screenshot of the constructed short-course home page, illustrating the typical integration of learning resources, activities, and supports. The segment of image on the right shows an example of a multimedia resource—in this case, a discussion on measuring quadriceps strength.



Outcomes

Kirkpatrick's hierarchy of educational outcomes proposes that training effects be examined for four levels of impact: (1) participant reaction, (2) participant knowledge, (3) participant change in behavior, and (4) change in health outcomes [36]. Outcomes were measured across levels of impact 1 to 3, with the same measures of both groups taken at the same time. Level 1 outcomes were measured through program attendance and adherence (assessed by signatures on a register of attendance for face-to-face delivery and computer-generated usage reports for Web-based participants), self-reported satisfaction (assessed via an electronic survey emailed to both groups), participants' ratings of the relevance of the program content to their current work roles, and self-reported estimates of time spent engaged with the learning resources. Level 2, knowledge, was measured via a 1-hour knowledge test, conducted approximately 1 week after completion of the program, and an assignment submission requiring a description of an exercise program tailored to a hypothetical client scenario. Level 3 outcomes were measured by self-reported change in practice, including a self-report of whether participants had changed their practices since completing the program. Change in practice was first measured through asking participants "Since completing the program, have you changed any aspects of how you manage your falls and balance clients?" and then, by open text comment, "If you answered 'yes' to the question above, please indicate in the space provided below in what way the program has changed your management of falls and balance clients." This question was then followed by a series of closed questions asking participants to provide their opinion about specific aspects of exercise prescription, such as "I attempt to use more motivational interviewing techniques." Participants responded to these closed questions using a Likert scale with five response categories (strongly agree, through to strongly disagree). Participants could not see these specific items until after they had completed the open-ended question so that these items would not influence the response to the open-ended question.

Procedure

Randomization was stratified by professional group (physiotherapist, occupational therapist, nurse, exercise physiologist) and nearest live program-delivery venue. An independent research assistant recruited participants in order of receipt of their registration and was unaware of group allocation. For the random allocation of participants, a computerized random number sequence was generated using permuted blocks of two, four, and six participants and stratified by geographical location of recruitment. Examination scores were automatically corrected using the online learning system that delivered the examination to both groups. Assignment submissions were de-identified before being forwarded to a blinded assessor.

XSL•FO
RenderX

Assignment submissions from both groups were scanned and reprinted to remove potential visual differences. The independent assessor, blind to the group allocation, assessed the assignments against pre-prepared assessment criteria. Thematic analysis of open text responses was verified by two researchers, with any lack of consensus referred to a third reviewer.

Statistical Analysis

Demographic characteristics of the two groups were compared at baseline using the chi-square test for items with binary data, and a 2-sample *t* test for comparing items with continuous data. Survey questions for measuring Kirkpatrick's hierarchy domains *participation reaction* and *change in behavior* were custom developed by the investigators based on a previously published description of these domains [36]. Responses to these items were subjected to principal-factor factor analysis with rotation. Four factors identified with eigenvalues greater than 1.0 were then reviewed for consistency and redundancy. Items were removed from factors if their rotated factor loading was less than 0.5 or the item uniqueness was less than 0.2. We then reviewed the factors generated for coherence and plausibility, which led to the removal of one factor, as the two items that it was composed of were seemingly unrelated. The remaining three factors were named based on the included items. These factors were (1) satisfaction with content and its relevance, (2) satisfaction with the support and facilitation, and (3) change in clinical behavior. Cronbach alpha for each factor was .92, .77, and .84, respectively. These factors were considered to be consistent with the theoretical framework used to originally guide item development. A simple summative score was then constructed for each factor following the principles of classical test theory. The resulting summative score represented the domain named by that factor. The wording of items contained in each factor is presented in Multimedia Appendix 1. The scale score range for each factor was 0–35, 0–15, and 0–30 respectively. The effect of education-delivery approach on each factor was examined by linear regression using the factor score as the dependent variable and a dummy variable coding for education approach (0 = face-to-face, 1 = Web-based) as the sole independent variable in the model.

Knowledge test and assignment scores were marked out of a score range of 0 (no correct answers) to 30 for the knowledge test and 100 for the assignment, before being converted to a percentage. These scores were examined for normality of distribution by examination of histograms and tests for skewness (skewness = –0.56, $P = .02$, indicating that the null hypothesis that the data is normally distributed can be rejected [37]), where a high proportion of scores were loaded toward the maximum score. We therefore used the nonparametric rank sum test (Wilcoxon test or Mann-Whitney *U* test) to compare groups on this outcome. The amount of time individuals spent engaged with optional further reading materials was also found to be skewed (skewness = 4.00, $P < .001$); hence, we also used a rank sum test for comparison between groups.

Responses to open-ended questions were analyzed thematically by two investigators. We used the statistics software package Stata version 11 (StatCorp LP, College Station, TX, USA) for all data analyses.

Results

Of the 166 participants initially screened, representing a range of professions including physiotherapy, occupational therapy, exercise physiology, and nursing, there was gradual attrition from randomization to participation in the trial (n = 67 Web-based, n = 68 face-to-face), to completion of the educational content (n = 44 Web-based, n = 50 face-to-face), to completion of the posteducation knowledge test (n = 43 Web-based, or 36% attrition from initial random allocation, and n = 49 face-to-face, or 28% attrition from random allocation; Figure 2).

Participant demographics are presented in Table 1 and indicate a relatively even distribution of baseline characteristics.

XSL•FO
RenderX

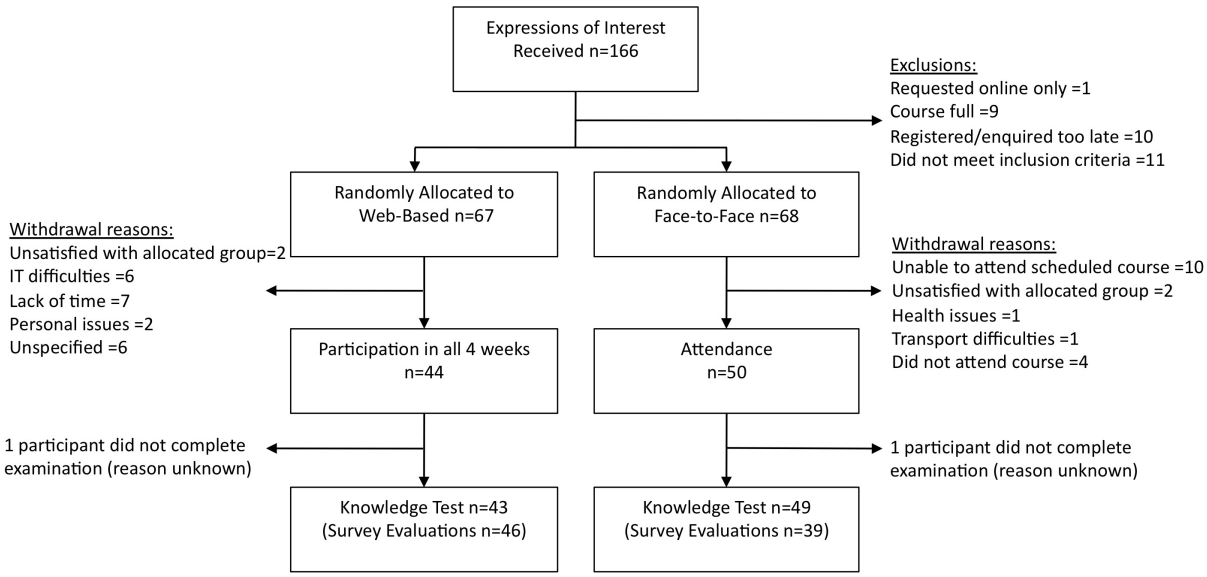
Table 1. Demographic characteristics of study participants by group

Demographic item	Education mode		*P* value
	Web-based	Face-to-face	
Gender (male), n (%)[a]	10 (22%)	7 (18%)	.43
Previous falls research participation, n (%)[a]	2 (4%)	5 (13%)	.18
Previous falls publication, n (%)[a]	1 (2%)	0 (0%)	.35
Previous falls professional development, n (%)[a]	11 (23%)	10 (25%)	.85
Profession, n (%)[a]			
Occupational therapy	5 (11%)	3 (8%)	.97
Physiotherapy	26 (57%)	20 (51%)	.93
Nursing	10 (22%)	11 (28%)	.95
Exercise physiology	4 (9%)	4 (9%)	.96
Years since qualification, mean (SD)[b]	4.17 (1.75)	4.15 (1.56)	.66

[a] Chi-square test.

[b] Two-sample *t* test.

Figure 2. CONSORT flowchart showing attrition of study participants (IT = information technology).



Kirkpatrick's Level 1: Reaction

Mean (SD) scores in the domain *satisfaction with content and its relevance* appeared to be similar for the two groups: Web-based, 25.73 (5.14), and face-to-face, 26.11 (5.41). Comparison between groups using linear regression for this domain did not identify a statistically significant difference: beta coefficient (95% confidence interval [CI]), –.38 (–2.70 to 1.94); *P* = .75. Mean (SD) scores in the *satisfaction with course facilitation and support* were also comparable between groups:

XSL•FO
RenderX

Web-based, 11.61 (2.00), and face-to-face, 12.08 (1.54). The difference was also not statistically significant: beta coefficient (95% CI), –.47 (–1.29 to 0.35); $P = .25$.

Web-based participants reported spending a mean (SD) of 8.7 (6.6) hours engaged with the compulsory learning resources as compared with the 7 hours required of the face-to-face group, which was controlled by the venue's opening hours. Web-based participants reported spending significantly more time engaged with the additional learning materials than the face-to-face group: median (interquartile range [IQR]), 1.0 (0.8–2.0) hours compared with 0.0 (0.0–1.0) hours; rank sum $P = .002$.

Kirkpatrick's Level 2: Knowledge Outcome

Knowledge test results were comparable between face-to-face and Web-based groups: median (IQR) for Web-based, 90.00 (70.89–90.67), and for face-to-face, 80.56 (70.67–90.00); rank sum $P = .07$. The median (IQR) scores for the exercise assignment were also comparable: Web-based, 78.6 (68.5–85.1), and face-to-face, 78.6 (70.8–86.9); rank sum $P = .61$.

Kirkpatrick's Level 3: Change in Practice

Mean (SD) scores in the domain *change in clinical behavior* were similar between groups: Web-based, 21.75 (4.40), and face-to-face, 21.88 (3.24). Comparison between groups using linear regression for this domain did not identify a statistically significant difference: beta coefficient (95% CI), –.13 (–1.99 to 1.74), $P = .89$.

Thematic analysis of the optional open text comments by participants revealed that Web-based participants were primarily disclosing changes in their application of motivational interviewing strategies (8/22), along with changes to improve the competency of their client assessments (8/22). In contrast, participants in the face-to-face mode of delivery did not comment regarding motivational interviewing (0/21), with change in assessment (10/21) and change in exercise prescription (12/21) as the two components of practice most frequently identified.

Of the Web-based participants, 24% (11/46) reported being apprehensive about undertaking a Web-based program before the program commenced, rating agree or strongly agree to the statement on a 5-point Likert scale ranging from strongly disagree to strongly agree. After the program had been completed, 70% (32/46) of respondents indicated that they quickly became accustomed to the Web-based environment and 80% (37/46) indicated agreement that they would be willing to undertake another Web-based program.

Discussion

Principal Results

Web-based and face-to-face approaches for providing training in the field of exercise prescription for falls prevention produced comparable results for all three levels of Kirkpatrick's hierarchy of educational outcomes. Previous studies have reported that Internet-based CPD can increase the acquisition of health professional knowledge and lead to change in clinical behavior in single-profession educational studies [28]. Previous head-to-head studies of Web-based versus live delivery of CPD have examined procedural and theoretical skills applied to decision making, such as recognizing the need for referral for further medical testing. This study provides evidence that outcomes of Web-based CPD are not significantly different from those of face-to-face CPD applied to an interprofessional audience and a subject matter that encompasses a broad range of practical skills including clinical decision making, hands-on skills, and high-level communication.

With comparable results between delivery methods, practical considerations should arguably drive the choice of delivery mode. These may include being able to create a more standardized educational product that is less influenced by the style, experience, and knowledge of individual presenters, and the ability to protect corporate knowledge with changes in personnel, or the educational institution's information and communication technology capabilities, resources, and support structure.

This study has demonstrated that, in the field of falls prevention, Web-based education may provide results in terms of Kirkpatrick's levels 1 to 3 equivalent to those of face-to-face education. This enhances the capacity for upgrading the skills of health professionals for whom geographic isolation is a barrier to participation in CPD in this area. With greater skills they may be able to better meet the needs of older adults at risk of falls residing in rural and remote areas.

Limitations and Future Directions

Some strengths and limitations may have affected the collection, analysis, and generalizability of the intervention results.

Limitations in data collection may have arisen from the custom design of the survey questions intended to measure the Kirkpatrick hierarchy domains. The examination was open book, although no overall grade was awarded to the student for the short course, so as to minimize the motivation for collusion and looking up answers from alternative sources. We were unable to control for collusion between participants in either the Web-based or face-to-face intervention group, as the examination was conducted online and off-site.

We noted moderate correlation between the knowledge test and assignment scores (Pearson $r = .44$); however, this is not surprising considering that the intent of the exam was to test memory of theoretical knowledge, and the assignment was to test applied clinic-based problem solving and ability in exercise design.

Supporting positive changes in Kirkpatrick's hierarchy level 1 may have included program design aspects aimed at decreasing the transactional distance, or feeling of isolation and separation from the learning group, through an actively engaged facilitator, available discussion forums, and multiple tasks encouraging participant networking. We provided numerous opportunities for feedback on knowledge and practical skills, through weekly knowledge tests with automated feedback, video-skill submissions with feedback, and selected peer submissions published through the online learning system, to allow active reflection and self-evaluation of skills mastery. Participants were additionally supported by an extensive information technology support network provided by the administering

XSL•FO
RenderX

university, with helplines available from 8:00 AM to 8:00 PM Monday to Friday. Factors that may have negatively affected satisfaction and attrition may have included nursing participants indicating that the program skills and objectives were not as relevant to their profession as to participants with physiotherapy qualifications. Information technology difficulties were also experienced as a direct result of complications arising from the process of randomly assigning participants to Web-based or face-to-face delivery and through upgrading of the program administrators' skills in the first iteration of the program. It is anticipated that many of these difficulties could be eliminated in future modeling of the program. The program was offered to participants at no cost, which decreases the participants' commitment to the program and negatively affects attrition [38]. The program design required a significantly high level of information technology literacy by the participants. Although clinicians' use of the Internet and other information technologies is increasing exponentially, with between 78% and 85% of clinicians accessing health information via the World Wide Web [39], universal access and competency in information and communication technologies cannot be assumed [11]. Contemporary university students, of generation Y, are generally computer literate and embrace new technology [40]. However, given that the participants in this trial were postgraduates, of generation X, with a mean of 3.9 years since obtaining their professional qualifications, we did not know whether participants would be able to readily source and manage information technology requirements, particularly with tasks involving electronic submission of self-videos. Although 70% of respondents reported that they quickly became accustomed to the Web-based environment, only 80% of Web-based participant respondents reported that they would complete another Web-based program.

The follow-up regarding clinical practice change was limited to a short-term follow-up. It is feasible that participants ceased using these skills in the medium or longer term, or conversely increased their use of these skills. It is unknown whether patients of participants had health outcomes that were improved to a greater or lesser extent as a result of the participants' educational approach. We excluded, but could have included, some health professional groups from the present study that could have a role in the clinical management of older people at risk of falls.

This study also has limitations in how broadly these results can be generalized into other CPD areas. The content focused on one clinical area, and it is possible that other subject areas may not be as readily adapted into a Web-based format (for example, teaching spinal manipulation techniques).

Extensions of this research could include investigations of cost minimization, factors influencing willingness to enroll in Web-based education, and economic efficiency of face-to-face and Web-based approaches to CPD. If Web-based education proved to be a more efficient method for increasing the number of skilled clinicians in this area, yet comparable in its educational outcomes as shown in this study, then this may influence public health strategy in addressing falls in our community, and have wider implications still on how CPD is delivered around other health priorities. Another extension of this research would be to assess the impact of program participation on health outcomes, addressing level 4 of Kirkpatrick's hierarchy.

Conclusions

Face-to-face and Web-based approaches to the delivery of education to clinicians on the subject of exercise prescription for falls prevention produced equivalent results in all of the outcome domains, with the exception of Web-based participants reporting more time engaged with the optional learning resources. Practical considerations should arguably drive the choice of delivery method, which may favor Web-based provision for its ability to overcome access issues for health professionals in regional and remote settings.

Acknowledgments

The authors would like to thank the Department of Health, Victoria, Australia, for providing funding for this research.

Conflicts of Interest

None declared

Multimedia Appendix 1

The wording of items contained in each factor.

[PDF File (Adobe PDF File), 52KB - jmir_v13i4e116_app1.pdf]

References

1. Shepard K, Jensen G. Handbook of Teaching for Physical Therapists. 2nd edition. Boston, MA: Butterworth-Heinemann; 2002.
2. Whitcomb ME. CME reform: an imperative for improving the quality of medical care. Acad Med 2002 Oct;77(10):943-944. [Medline: 12377664]
3. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. JAMA 1995 Sep 6;274(9):700-705. [Medline: 7650822]

XSL•FO
RenderX

4. Davis DA, Taylor-Vaisey A. Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. CMAJ 1997 Aug 15;157(4):408-416 [FREE Full text] [Medline: 9275952]
5. Mazmanian PE, Davis DA. Continuing medical education and the physician as a learner: guide to the evidence. JAMA 2002 Sep 4;288(9):1057-1060. [Medline: 12204068]
6. Schoo AM, Stagnitti KE, Mercer C, Dunbar J. A conceptual model for recruitment and retention: allied health workforce enhancement in Western Victoria, Australia. Rural Remote Health 2005 Dec;5(4):477 [FREE Full text] [Medline: 16375575]
7. Curran VR, Fleet L, Kirby F. Factors influencing rural health care professionals' access to continuing professional education. Aust J Rural Health 2006 Apr;14(2):51-55. [doi: 10.1111/j.1440-1584.2006.00763.x] [Medline: 16512789]
8. Randolph DS. Predicting the effect of extrinsic and intrinsic job satisfaction factors on recruitment and retention of rehabilitation professionals. J Healthc Manag 2005 Feb;50(1):49-60. [Medline: 15729907]
9. Rabinowitz HK, Diamond JJ, Markham FW, Paynter NP. Critical factors for designing programs to increase the supply and retention of rural primary care physicians. JAMA 2001 Sep 5;286(9):1041-1048 [FREE Full text] [Medline: 11559288]
10. Biggs J. What the student does: teaching for enhanced learning. Higher Educ Res Dev 1999;18(1):57-75. [doi: 10.1080/0729436990180105]
11. Coffee J, Hillier S. Teaching pre-cursor clinical skills using an online audio-visual tool: an evaluation using student sesponses. J Online Learn Teach 2008;4(4):469-476.
12. Salyers VL. Teaching psychomotor skills to beginning nursing students using a web-enhanced approach: a quasi-experimental study. Int J Nurs Educ Scholarsh 2007;4:Article11 [FREE Full text] [doi: 10.2202/1548-923X.1373] [Medline: 17402927]
13. Barker SP. Comparison of effectiveness of interactive videodisc versus lecture-demonstration instruction. Phys Ther 1988 May;68(5):699-703 [FREE Full text] [Medline: 3283787]
14. Coffee J. Using an on line audio visual aid to facilitate the teaching and learning of clinical reasoning. Focus Health Prof Educ 2007;9(3):89-91.
15. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol 1999 Dec;77(6):1121-1134. [Medline: 10626367]
16. Miller PA. Self-Assessment: The Disconnect between Research and Rhetoric. Physiother Can 2008;60(2):117-124. [doi: 10.3138/physio.60.2.117] [Medline: 20145775]
17. Kronz JD, Silberman MA, Allsbrook WC, Epstein JI. A web-based tutorial improves practicing pathologists' Gleason grading of images of prostate carcinoma specimens obtained by needle biopsy: validation of a new medical education paradigm. Cancer 2000 Oct 15;89(8):1818-1823. [Medline: 11042578]
18. Harris JM, Salasche SJ, Harris RB. Can Internet-based continuing medical education improve physicians' skin cancer knowledge and skills? J Gen Intern Med 2001 Jan;16(1):50-56. [Medline: 11251750]
19. Pagnanelli G, Soyer HP, Argenziano G, Talamini R, Barbati R, Bianchi L, et al. Diagnosis of pigmented skin lesions by dermoscopy: web-based training improves diagnostic performance of non-experts. Br J Dermatol 2003 Apr;148(4):698-702. [Medline: 12752126]
20. Curran VR, Hoekman T, Gulliver W, Landells I, Hatcher L. Web-based continuing medical education (I): field test of a hybrid computer-mediated instructional delivery system. J Contin Educ Health Prof 2000;20(2):97-105. [doi: 10.1002/chp.1340200206] [Medline: 11232226]
21. Curran VR, Hoekman T, Gulliver W, Landells I, Hatcher L. Web-based continuing medical education. (II): Evaluation study of computer-mediated continuing medical education. J Contin Educ Health Prof 2000;20(2):106-119. [doi: 10.1002/chp.1340200207] [Medline: 11232219]
22. Hinkka H, Kosunen E, Metsänoja R, Lammi UK, Kellokumpu-Lehtinen P. General practitioners' attitudes and ethical decisions in end-of-life care after a year of interactive Internet-based training. J Cancer Educ 2002;17(1):12-18. [doi: 10.1080/08858190209528785] [Medline: 12000099]
23. Harris JM, Kutob RM, Surprenant ZJ, Maiuro RD, Delate TA. Can Internet-based education improve physician confidence in dealing with domestic violence? Fam Med 2002 Apr;34(4):287-292. [Medline: 12017143]
24. Kemper KJ, Amata-Kynvi A, Sanghavi D, Whelan JS, Dvorkin L, Woolf A, et al. Randomized trial of an internet curriculum on herbs and other dietary supplements for health care professionals. Acad Med 2002 Sep;77(9):882-889. [Medline: 12228083]
25. Gerbert B, Bronstone A, Maurer T, Berger T, McPhee SJ, Caspers N. The effectiveness of an Internet-based tutorial in improving primary care physicians' skin cancer triage skills. J Cancer Educ 2002;17(1):7-11. [doi: 10.1080/08858190209528784] [Medline: 12000111]
26. Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? JAMA 1999 Sep 1;282(9):867-874 [FREE Full text] [Medline: 10478694]
27. Khatony A, Nayery ND, Ahmadi F, Haghani H, Vehvilainen-Julkunen K. The effectiveness of web-based and face-to-face continuing education methods on nurses' knowledge about AIDS: a comparative study. BMC Med Educ 2009;9:41 [FREE Full text] [doi: 10.1186/1472-6920-9-41] [Medline: 19591678]

XSL•FO
RenderX

28. Fordis M, King JE, Ballantyne CM, Jones PH, Schneider KH, Spann SJ, et al. Comparison of the instructional efficacy of Internet-based CME with live interactive CME workshops: a randomized controlled trial. JAMA 2005 Sep 7;294(9):1043-1051 [FREE Full text] [doi: 10.1001/jama.294.9.1043] [Medline: 16145024]
29. Pointer S, Harrison J, Bradley C. Australian Institute of Health and Welfare. 2003 Jul. National Injury Prevention Plan Priorities for 2004 and beyond: Discussion Paper URL: http://www.nisu.flinders.edu.au/pubs/reports/2003/injcat55.pdf [accessed 2011-11-07] [WebCite Cache ID 631dFYSqT]
30. Cassell E, Clapperton A. Preventing injury in Victorian seniors aged 65 years and older. Hazard 2008;67:1-24.
31. Moller J. Commonwealth Department of Health and Ageing. 2003 Jul. Projected costs of fall related injury to older persons due to demographic change in Australia URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/1063D7B1A50DA9B5CA256F1900040484/$File/falls_costs.pdf [accessed 2011-09-28] [WebCite Cache ID 622PUjYZx]
32. Gillespie LD, Gillespie WJ, Robertson MC, Lamb SE, Cumming RG, Rowe BH. Interventions for preventing falls in elderly people. Cochrane Database Syst Rev 2003(4):CD000340. [doi: 10.1002/14651858.CD000340] [Medline: 14583918]
33. Sherrington C, Whitney JC, Lord SR, Herbert RD, Cumming RG, Close JC. Effective exercise for the prevention of falls: a systematic review and meta-analysis. J Am Geriatr Soc 2008 Dec;56(12):2234-2243. [doi: 10.1111/j.1532-5415.2008.02014.x] [Medline: 19093923]
34. Thomas S, Mackintosh S, Halbert J. Does the 'Otago exercise programme' reduce mortality and falls in older adults?: a systematic review and meta-analysis. Age Ageing 2010 Nov;39(6):681-687 [FREE Full text] [doi: 10.1093/ageing/afq102] [Medline: 20817938]
35. Day L, Finch CF, Harrison JE, Hoareau E, Segal L, Ullah S. Modelling the population-level impact of tai-chi on falls and fall-related injury among community-dwelling older people. Inj Prev 2010 Oct;16(5):321-326. [doi: 10.1136/ip.2009.025452] [Medline: 20643871]
36. Hutchinson L. Evaluating and researching the effectiveness of educational interventions. BMJ 1999 May 8;318(7193):1267-1269 [FREE Full text] [Medline: 10231262]
37. D'Agostino RB, Belanger AJ, D'Agostino RBJ. A suggestion for using powerful and informative tests of normality. Am Stat 1990;44:316-321.
38. Kazmirski G. Marketing quality and value to the managed care market. Top Health Inf Manage 1998 Nov;19(2):62-69. [Medline: 10338715]
39. Bennett NL, Casebeer LL, Kristofco RE, Strasser SM. Physicians' Internet information-seeking behaviors. J Contin Educ Health Prof 2004;24(1):31-38. [doi: 10.1002/chp.1340240106] [Medline: 15069910]
40. Oblinger DG. Space as a change agent. In: Oblinger DG, editor. Learning Spaces. Washington, DC: Educause; 2006:1.1-1.4.

Abbreviations

CPD: continuing professional development
IQR: interquartile range

Edited by G Eysenbach; submitted 04.11.10; peer-reviewed by O Mashrabi, S Fraser; comments to author 15.02.11; revised version received 13.05.11; accepted 25.09.11; published 22.12.11.

Please cite as:
Maloney S, Haas R, Keating JL, Molloy E, Jolly B, Sims J, Morgan P, Haines T
Effectiveness of Web-Based Versus Face-To-Face Delivery of Education in Prescription of Falls-Prevention Exercise to Health Professionals: Randomized Trial
J Med Internet Res 2011;13(4):e116
URL: http://www.jmir.org/2011/4/e116/
doi:10.2196/jmir.1680
PMID:22189410

©Stephen Maloney, Romi Haas, Jennifer L Keating, Elizabeth Molloy, Brian Jolly, Jane Sims, Prue Morgan, Terry Haines. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Hospital-Based Nurses' Perceptions of the Adoption of Web 2.0 Tools for Knowledge Sharing, Learning, Social Interaction and the Production of Collective Intelligence

Adela S.M Lau[1], PhD

Interdisciplinary Programs Office, The Hong Kong University of Science and Technology, HKUST, Hong Kong, China

Corresponding Author:
Adela S.M Lau, PhD
Interdisciplinary Programs Office
The Hong Kong University of Science and Technology
HKUST
Clear Water Bay
Kowloon
Hong Kong
China
Phone: 852 23585729
Fax: 852 31040026
Email: adela@ust.hk

Abstract

Background: Web 2.0 provides a platform or a set of tools such as blogs, wikis, really simple syndication (RSS), podcasts, tags, social bookmarks, and social networking software for knowledge sharing, learning, social interaction, and the production of collective intelligence in a virtual environment. Web 2.0 is also becoming increasingly popular in e-learning and e-social communities.

Objectives: The objectives were to investigate how Web 2.0 tools can be applied for knowledge sharing, learning, social interaction, and the production of collective intelligence in the nursing domain and to investigate what behavioral perceptions are involved in the adoption of Web 2.0 tools by nurses.

Methods: The decomposed technology acceptance model was applied to construct the research model on which the hypotheses were based. A questionnaire was developed based on the model and data from nurses (n = 388) were collected from late January 2009 until April 30, 2009. Pearson's correlation analysis and *t* tests were used for data analysis.

Results: Intention toward using Web 2.0 tools was positively correlated with usage behavior (r = .60, P < .05). Behavioral intention was positively correlated with attitude (r = .72, P < .05), perceived behavioral control (r = .58, P < .05), and subjective norm (r = .45, P < .05). In their decomposed constructs, perceived usefulness (r = .7, P < .05), relative advantage (r = .64, P < .05), and compatibility (r = .60,P < .05) were positively correlated with attitude, but perceived ease of use was not significantly correlated (r = .004, P < .05) with it. Peer (r = .47, P < .05), senior management (r = .24,P < .05), and hospital (r = .45, P < .05) influences had positive correlations with subjective norm. Resource (r = .41,P < .05) and technological (r = .69,P < .05) conditions were positively correlated with perceived behavioral control.

Conclusions: The identified behavioral perceptions may further health policy makers' understanding of nurses' concerns regarding and barriers to the adoption of Web 2.0 tools and enable them to better plan the strategy of implementation of Web 2.0 tools for knowledge sharing, learning, social interaction, and the production of collective intelligence.

(J Med Internet Res 2011;13(4):e92) doi:10.2196/jmir.1398

KEYWORDS

E-learning & Collective Intelligence; Web 2.0 tools; human behavioral adoption

XSL•FO
RenderX

Introduction

Web 2.0 tools are people-based knowledge sharing, learning, social interaction, and collective intelligence tools that support knowledge collaboration, exchange, sharing, and creation. They provide the platform and tools such as blogs, wikis, podcasts, social bookmarks, really simple syndication (RSS), tags, and social networking software to enable learners to interact and communicate in a virtual environment [1,2]. Following the rapid growth in usage of Web 2.0 tools in knowledge sharing, learning, social interaction, and the production of collective intelligence [1,3-7], this paper aimed to investigate how Web 2.0 tools are to be applied in the nursing domain for these purposes and to investigate the behavioral perceptions of the adoption of Web 2.0 tools by nurses. The objectives of this study were to investigate how Web 2.0 tools can be applied for knowledge sharing, learning, social interaction, and the production of collective intelligence; to design a research model to identify factors influencing nurses' intention to adopt the tool; to design hypotheses and a questionnaire based on the model; and to collect the data and identify the factors influencing nurses' intention to adopt Web 2.0 tools for knowledge sharing, learning, social interaction, and the production of collective intelligence.

In the following sections, the use of Web 2.0 tools for the purposes mentioned above is discussed. The human adoption behavior models are reviewed and the proposed model, hypotheses, and questionnaire are designed. The sampling and statistical techniques used are also presented as well as the pilot testing and data collection results. Finally, the implications of the results and conclusions are discussed.

How Web 2.0 Tools Support Knowledge Sharing, Learning, Social Interaction, and the Production of Collective Intelligence by Nurses

Web 2.0 technologies such as blogs, wikis, really simple syndication, podcasts, tags, social bookmarks, and social networking software have the features of social interaction and collaboration to facilitate knowledge sharing, learning, social interaction, and the production of collective intelligence over the Internet [8,9]. Web 2.0 technologies allow a community to publish and edit a document collaboratively in a virtual environment [10]. Through such social interaction and collective intelligence, knowledge is created, exchanged, and shared.

Blog

A blog is a user-friendly content management tool that allows users (bloggers) to publish their own content on the Web [1,2,11-14]. A blogger shares his or her writings (blogs), gains comments or opinions from other bloggers, and links his or her blog to other blogs. Through such blog sharing and linkage, communities with the same interests and discussion topics are formed. Using blogs, nurses can learn about workplace experiences from each other, helping them to gain nursing knowledge from the virtual community and via social interaction [8].

Wiki

A wiki is a collaborative editing tool that allows authors to coedit a document [2,10-12,14,15]. A wiki has the features of content management, versioning control, rights management, and so forth [8]. Authors collaboratively edit, review, and revise a single document. Through such collaboration and collective intelligence, knowledge is created and acquired. Using a wiki, nurses can go through collaborative and reflective learning processes to gain knowledge from other nurses and apply this knowledge to solving a problem.

Really Simple Syndication

Really simple syndication (RSS) is a feed reader for content distribution, dissemination, and acquisition over Internet sources [14]. The RSS feed reader automatically sends an alert signal and pushes the updated content to RSS subscribers so that they can gather the most up-to-date information in real time. Using the RSS, nurses can share Internet resources with others to facilitate knowledge sharing, updating, and learning in a real-time environment.

Podcast

A podcast is a series of audio or video digital media files for playback on portable media players and computers [1,11,16]. It can be syndicated, subscribed to, and downloaded automatically when the content is updated. Podcasters distribute and disseminate digital media files over the Internet, and subscribers can obtain podcasts via an RSS feed reader at any time [8]. Using RSS, nurses can share or capture nursing skills and techniques in image, audio, or video files with other nurses via RSS to enable nursing learning and production of collective intelligence to take place anytime and anywhere.

Tags

Tags are the keywords or terms for describing digital media content such as social bookmarks, audio clips, video clips, blogs, wikis, and websites. Tags are built by a community and are used to describe its content [8]. The tag cloud function collects and counts the number of tags used by a community and groups and classifies them into different topics that enable a search engine to search more accurately [17]. Nurses can tag websites or learning resources for sharing.

Social Bookmark

A social bookmark enables Internet users to store, organize, search, and manage webpage bookmarks [2,17] and is described by tags. By clustering the bookmark's tags, bookmark pages can be linked and clustered into different topics. Nurses can use social bookmarks for knowledge sharing and learning, to shorten their resource searching costs, and to facilitate the social learning atmosphere by sharing resources.

Social Networking Software

Social networking software typically provides social networking functions such as audio/video conferencing, Internet protocol (IP) telephony, desktop sharing, chat rooms, and whiteboards to enable a community to communicate and interact in a virtual environment. Professional social networking software may provide community-building functions such as an electronic

XSL•FO
RenderX

portfolio, resume builder, and social networking so that people can be connected together to form online communities to exchange and share knowledge [8]. Using social networking software, nurses can build and maintain their social community, thereby facilitating social interaction, learning, and the production of collective intelligence over the Internet, in a similar way as patients are doing [18,19].

In summary, Web 2.0 tools provide the features of collaborative work, social networking, community, and self-management. By using social networking software, blogs, and wikis, nurses can build communities and learn through knowledge collaboration, exchange, and sharing [7]. Web 2.0 tools provide a networked environment for learners to interact with each other in a single place and to learn new knowledge through social interaction and reflective learning processes. RSS, podcasts, tags, and social bookmarks are some other Web 2.0 tools that link up Internet learning resources in a virtual, distributed, and real-time environment that facilitates knowledge sharing and learning. However, the attitude of nurses to the adoption of Web 2.0 tools is critical to the success of its application for knowledge sharing, learning, and social interaction. What are the behavioral perceptions influencing nurses in the adoption of Web 2.0 tools? Since hospital-based nurses may require more collaboration, interaction, and knowledge sharing on patient care and nursing diagnosis than non–hospital-based nurses, this study mainly focuses on surveying hospital-based nurses.

Human Behavior Models

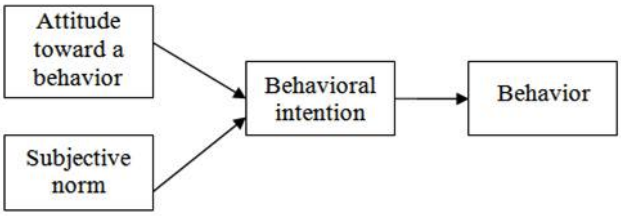
The theory of reasoned action (TRA), the theory of planned behavior (TPB) [20] and the technology acceptance model (TAM) [21,22] are the most widely used human behavior models [23,24,25] for studying human perceptions of the adoption of behaviors. The TRA (see Figure 1) predicts and explains the causes of behavior by evaluating a person's attitude and subjective norms [26,27]. The TPB (see Figure 2) is similar to what is advocated by the theory of reasoned action (TRA) but with the injection of perceived behavioral control [28] in which personal beliefs such as resources, opportunities, and obstacles are considered. In other words, the TPB studies not only the perceptions of social individual variables but also internal and external constraints on the behavior.

However, human behavior with regard to adoption of information technology (IT) cannot be described by these social individual variables and constraints alone. Human behavior may involve some practical concerns or facilitating conditions. Thus, the TAM (see Figure 3) was developed by Davis to explain computer usage behavior [29,30] and is more oriented to analyzing human behavior with regard to IT than the TRA and TPB [31]. The two attributes, perceived usefulness and ease of use [30], determine major external variables that may affect the human decision to use IT. In turn, they form the actual outcome of an action. However, subjective norm is abandoned in this model due to "its uncertain theoretical and psychometric status" [29]. In addition, perceived behavioral control is also omitted from the TAM.

Thus, the decomposed theory of planned behavior (DTPB) [32-34] is derived from the basic beliefs and structure of the theory of planned behavior model. In the DTPB model (see Figure 4), attitude, subjective norm, and perceived behavior control are further decomposed into smaller constructs. This provides a more comprehensive explanation of adoption behavior. It has been said that "the model becomes more managerially relevant, pointing to specific factors that may influence adoption and usage." This DTPB model also takes advantage of TAM, as it identifies specific salient beliefs that may influence IT usage. It incorporates significant subfactors, including relative advantage, compatibility, normative influence (subjective norm), efficacy, and facilitating condition, which are important determinants of human behavior.

The DTBP is more managerially relevant—pointing to specific factors that may influence adoption and usage—and is more understandable as a result of focusing on specific factors of the technology acceptance research context. Thus, the DTBP was used as the framework of the research model to study nurses' behavioral perceptions on Web 2.0 tools adoption. Details of the proposed model are discussed in the following sections.

Figure 1. Theory of reasoned action.



XSL•FO
RenderX

Figure 2. Theory of planned behavior.

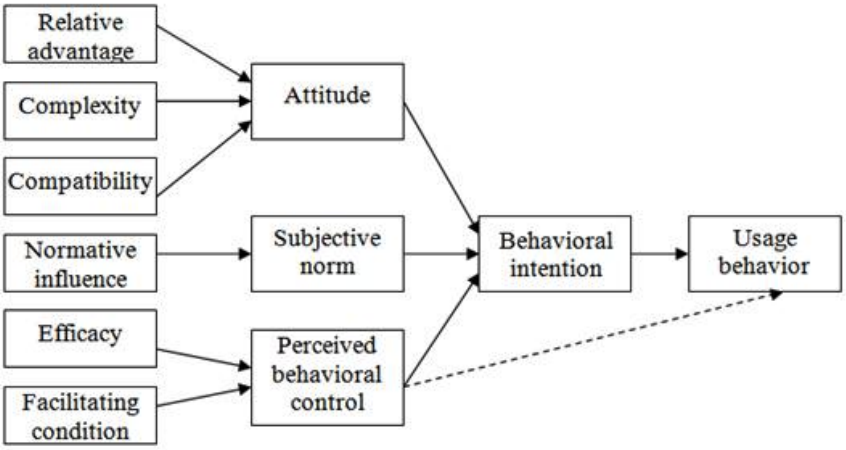
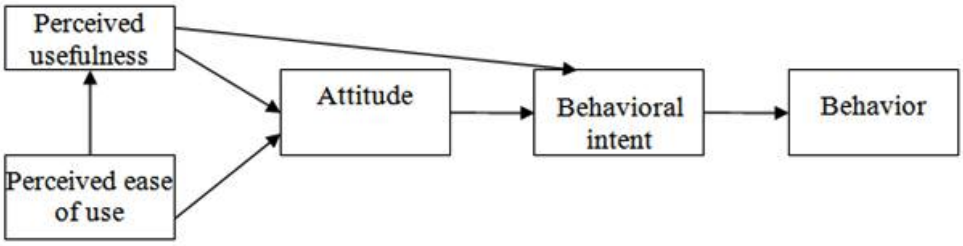
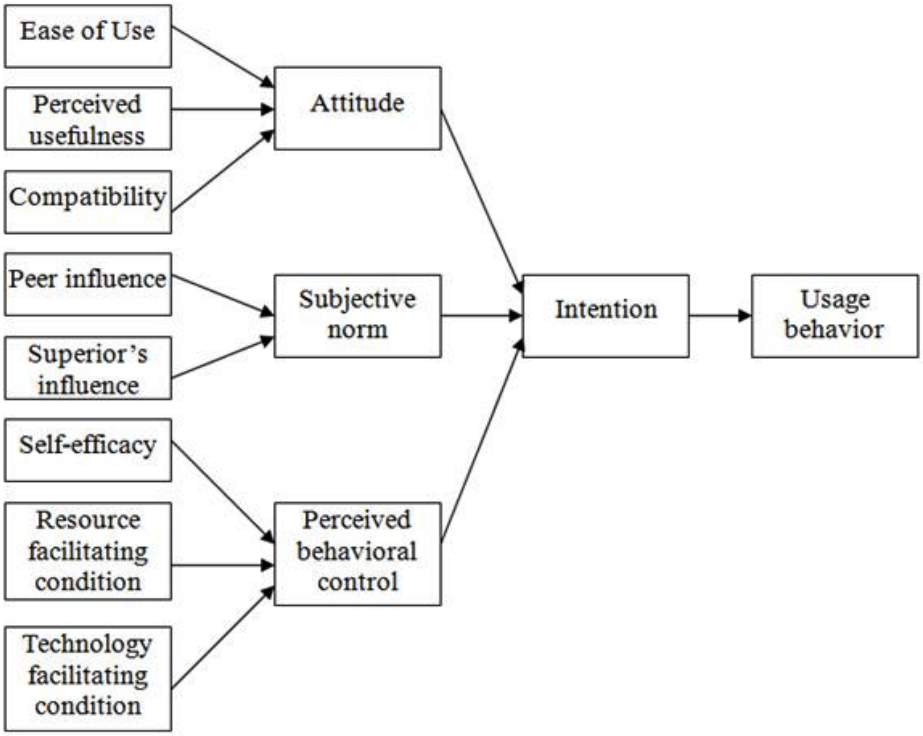


Figure 3. Technology acceptance model.



XSL•FO
RenderX

Figure 4. Decomposed theory of planned behavior.



Proposed Model

Based on the DTPB model, a new proposed theoretical framework was established for studying the adoption of Web 2.0 tools among Hong Kong nurses. The proposed model and model description (see Figure 5 and Table 1) for studying factors influencing the adoption of Web 2.0 tools among Hong Kong nurses are demonstrated below. The usage behavior of adopting Web 2.0 tools is determined by behavioral intention, and the three major determinants—attitude, subjective norm, and perceived behavioral control—are used to determine the behavioral intention. The three major determinants are further decomposed into detailed belief constructs. Perceived usefulness, perceived ease of use, relative advantage, and compatibility are the constructs that determine attitude to Web 2.0 tools. Peers, senior management, and hospital influences are the constructs that determine subjective norm. Resources and technology-facilitating conditions are the constructs that determine perceived behavioral control.

Based on the model, the hypotheses were set (see Table 2) and the questionnaire developed (see Multimedia Appendix 1).

XSL•FO
RenderX

Table 1. Model description

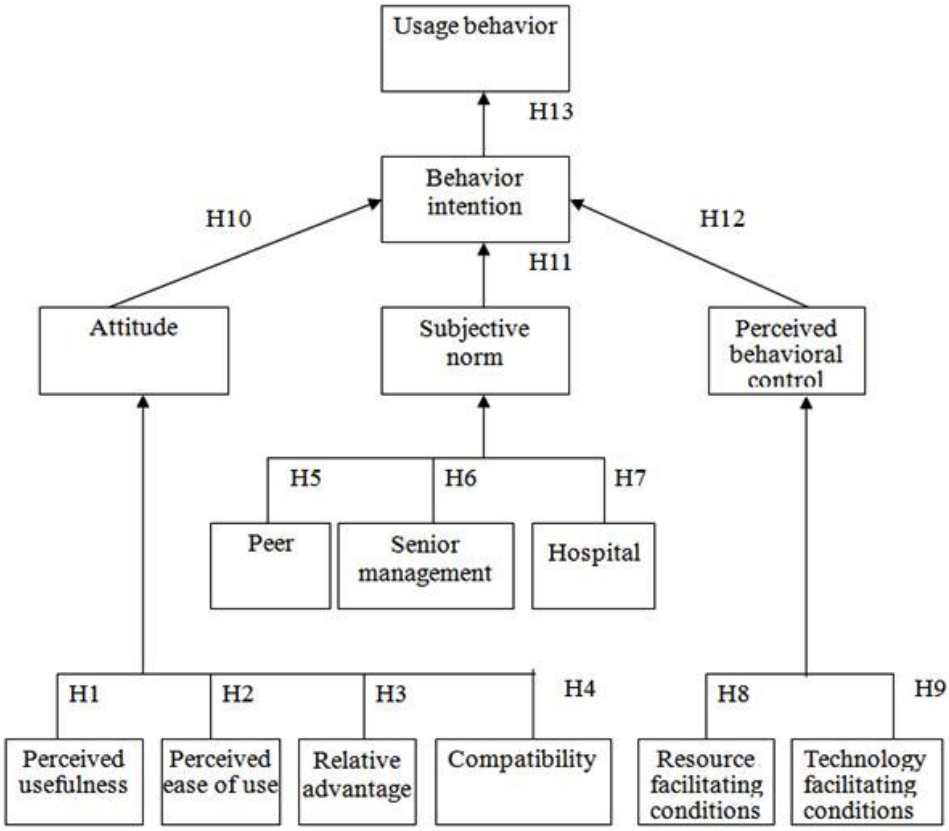
Construct	Description
Behavior or usage behavior (UB)	A person's performance of a specific action or an individual's decision to use Web 2.0 tools
Behavioral intention	A measure of the strength of intention to perform a specific action
Attitude	Whether a person possesses positive or negative feelings toward the behavior he or she performs
Perceived usefulness	The degree to which a person believes that using a particular system would enhance his or her job performance
Compatibility	The degree to which the innovation fits with the potential adopter's existing values, previous experiences, and current needs
Perceived ease of use	The degree to which a person believes that using Web 2.0 tools will be free of effort
Relative advantage	The degree to which an innovation is perceived as better than the idea it supersedes
Subjective norm	The perceived social pressure to perform a behavior
Peer influence, senior management influence, hospital influence	Influence of significant referents in our case
Perceived behavior control	The perception of the availability of skills, resources, and opportunities
Resource facilitating conditions	Resource factors such as time, money, and other factors relating to technology compatibility issues
Technology facilitating conditions	Available technology that is needed to make use of Web 2.0 tools

Table 2. Hypothesis setting

Hypothesis Number	Statement of Hypothesis	Question Number(s)
H1	Perceived usefulness of using Web 2.0 tools is positively correlated with attitude toward its adoption.	10, 11
H2	Perceived ease of use of Web 2.0 tools is positively correlated with attitude toward its adoption.	1, 2, 3
H3	Relative advantage of using Web 2.0 tools is positively correlated with attitude toward its adoption.	4, 5, 8
H4	Compatibility of using Web 2.0 tools is positively correlated with attitude toward its adoption.	6, 7, 9
H5	Peers' attitude toward using Web 2.0 tools is positively correlated with subjective norm.	16
H6	Senior management's attitude toward using Web 2.0 tools is positively correlated with subjective norm.	15
H7	Hospital's attitude toward using Web 2.0 tools is positively correlated with subjective norm.	17
H8	Resource facilitating conditions of Web 2.0 tools are positively correlated with perceived behavioral control.	18
H9	Technology facilitation conditions for using Web 2.0 tools are positively correlated with perceived behavioral control.	20
H10	Attitude toward Web 2.0 tools adoption is positively correlated with behavioral intention.	14
H11	Subjective norm concerning Web 2.0 tools adoption is positively correlated with behavioral intention.	19
H12	Perceived behavioral control of Web 2.0 tools adoption is positively correlated with behavioral intention.	21
H13	Behavioral intention of Web 2.0 tools adoption is positively correlated with usage behavior.	12, 13

XSL•FO
RenderX

Figure 5. Proposed model for studying factors influencing the adoption of Web 2.0 tools.



Methods

Sampling

There were 19,068 qualified nursing staff members in public hospitals during the fiscal year 2007-2008. With a confidence level of 95% and a confidence interval of 5, under the 50% preference, the required sample size was 377. Full-time qualified frontline registered nurses who were working under private and public hospitals and providing nursing care were included in this study. Enrolled and registered nurses who were working in outpatient departments, daycare centers, and the operating theater were excluded from this study.

Ethics Approval and Data Access

Ethics approval was obtained from the Research Approval Committee of the Hong Kong Polytechnic University. Data access in this study was approved by the nursing research approval committees of the Caritas Medical Centre, the School of Nursing at the Hong Kong Polytechnic University, and the Nethersole School of Nursing at the Chinese University of Hong Kong from late January 2009 until April 2009 when random sampling of qualified subjects was performed.

Data Collection Procedures

The purpose, nature, benefits, and risks of the study and the data collection procedures were explained to the subjects. Consent was obtained from all subjects involved in this survey.

The questionnaires were distributed and collected by the general nursing manager of the hospital. The return of the questionnaire was on a voluntary basis in a sealed envelope so as to ensure anonymity and confidentiality as stated in the cover letter of each questionnaire. The questionnaires for the sample recruited at the Hong Kong Polytechnic University and the Chinese University of Hong Kong were distributed by email and in person. The return of the questionnaire was also on a voluntary basis, with consents given by subjects and anonymity and data confidentiality being similarly ensured.

Statistical Analysis Methods

Pearson's correlation coefficient, *r*, and *t* test were used [35]. The correlation coefficient was used to study the strength of relationship between two constructs and the *t* test was used to determine whether the correlation itself was due to chance or not (ie, the significance level of the correlation).

Results

Pilot Testing

As a pilot test, the draft questionnaire was distributed to 34 nurses in the Hong Kong Polytechnic University, and 30 nurses returned them. The sample subjects found all the questions to be clear and understandable. No revisions were required. Only the reason for having nurses adopt Web 2.0 tools was found to be unclear. This aspect was then modified.

Response Rate of Sample

To meet the calculated sample size of 377, a total of 1053 questionnaires were distributed, and 392 questionnaires were

XSL•FO
RenderX

returned. Of these, 4 had not been completed, leaving 388 questionnaires for analysis for a response rate of 37%.

Demographic Characteristics of the Sample

The demographic characteristics of the sample are presented below (Table 3). The average age of the respondents was young, with the majority less than 30 years of age. Of the 388 respondents, 56% (219) were 21 to 30 years of age, 29% (111) were 31 to 40 years of age, and 14% (53) were 41 to 50 years of age. Only 1% (5) of respondents were 51 to 60 years of age, and none was over 60 years of age. Also, of the 388 respondents, 81% (314) were female and 19% (74) were male, while 66% (256) were single, 33% (129) were married, and 1% (3) were divorced.

In terms of education level, 26% (101/388) were subdegree holders (diploma), 64% (248/388) were degree holders, and 10% (39/388) had received a master's level education. Most respondents were receiving continuous education (69% or 266/388) and clinical training (56% or 216/388). Almost all respondents were in good physical health, defined as having no or only one medical problem (89% or 344/388). Again, of the 388 respondents, 91% (354) were registered nurses and 9% (34) were advanced practice nurses. The majority of the 388 respondents had more than 5 years' working experience (56% or 216).

Table 3. Demographic characteristics of the sample (n = 388)

Characteristics		Frequency (%)
Age group	21-30	219 (56%)
	31-40	111 (29%)
	41-50	53 (14%)
	51-60	5 (1%)
	>60	0 (0%)
Gender	Female	314 (81%)
	Male	74 (19%)
Marital status	Single	256 (66%)
	Married	129 (33%)
	Divorced	3 (1%)
Educational level	Sub-degree	101 (26%)
	Bachelor's	248 (64%)
	Master's	39 (10%)
Continuous education	No	122 (31%)
	Yes	266 (69%)
Clinical training	No	172 (44%)
	Yes	216 (56%)
Medical problems	0-1	344 (89%)
	2-3	36 (9%)
	>3	8 (2%)
Rank	Enrolled nurse	0 (0%)
	Registered nurse	354 (91%)
	Advanced practice nurse	34 (9%)
	Nursing officer	0 (0%)
Years of experience in nursing	<2	98 (0%)
	2-5	74 (19%)
	>5	216 (56%)

Survey Results and Implications

All of the correlation coefficients of the hypotheses were significant ($P < .05$), except for hypothesis 2 ($r = .004$, $P < .05$) (see Table 4). This implies that perceived ease of use of the Web 2.0 tools was not significant in predicting attitude toward their adoption. Therefore, hypothesis number H2 was rejected.

XSL•FO
RenderX

Table 4. Hypothesis testing results

Hypothesis	Content	Correlation Coefficient (Critical Value r = .08, P < .05)	Significance (Critical Value t_{386} = 1.65, P < .05)	Results
H1	Perceived usefulness of Web 2.0 tools is positively correlated with attitude toward its adoption.	.69	18.62	Accepted
H2	Perceived ease of use of Web 2.0 tools is positively correlated with attitude toward its adoption.	.004	Nil	Not Accepted
H3	Relative advantage of using Web 2.0 tools is positively correlated with attitude toward its adoption.	.64	16.53	Accepted
H4	Compatibility of using Web 2.0 tools is positively correlated with attitude toward its adoption.	.59	14.46	Accepted
H5	Peers' attitude toward using Web 2.0 tools is positively correlated with subjective norm.	0.47	10.31	Accepted
H6	Senior management's attitude toward using Web 2.0 tools is positively correlated with subjective norm.	0.24	4.835	Accepted
H7	Company's attitude toward using Web 2.0 tools is positively correlated with subjective norm.	0.45	9.95	Accepted
H8	Resource facilitating conditions of Web 2.0 tools are positively correlated with perceived behavioral control.	.41	8.85	Accepted
H9	Technology facilitation conditions for using Web 2.0 tools are positively correlated with perceived behavioral control.	.69	18.78	Accepted
H10	Attitude toward Web 2.0 tools adoption is positively correlated with behavioral intention.	.72	20.20	Accepted
H11	Subjective norm concerning Web 2.0 tools adoption is positively correlated with behavioral intention.	.45	9.81	Accepted
H12	Perceived behavioral control to Web 2.0 tools adoption is positively correlated with behavioral intention.	.58	14.02	Accepted
H13	Behavioral intention toward Web 2.0 tools adoption is positively correlated with usage behavior.	0.60	14.77	Accepted

Discussion

The first set of hypotheses showed that perceived usefulness (r = .69, P<.05), relative advantage (r = .64, P < .05), and compatibility (r = .59, P < .05) are positively correlated with attitude. The significance of the correlations between attitude and perceived usefulness (t = 18.62, P < .05), relative advantage (t = 16.53, P < .05), and compatibility (t = 14.46, P < .05) are high. This is because adopting Web 2.0 tools is not an objective decision but depends on how beneficial and useful [36,37] these tools will be to the nurses. In addition, the compatibility of Web 2.0 tools is also important to changing the actual behavior of nurses because of nurses' concerns regarding whether the virtual environment of Web 2.0 tools can support knowledge sharing, learning, and social interaction in the traditional way. However, perceived ease of use is not a concern since most individuals have experience using Web 2.0 tools such as blogs and RSS or have used Internet technology in wired or wireless environments via personal desktops, notebooks, shopping kiosks, or mobiles.

The testing of the second set of hypotheses revealed that peer (r = .47, P <.05), senior management (r = .24,P < .05), and hospital influences (r = .45, P < .05) are positively correlated with subjective norm. The significance of the correlations between subjective norm and peer (t = 10.31, P < .05), senior management (t = 4.83, P < .05), and hospital (t = 9.95, P < .05) influences and are high. Peer and hospital influences are more significant than senior management influence. This can be explained by the fact that Web 2.0 is a virtual environment for the community, and a virtual community cannot be formed without peer participation. Thus, peer participation in activities over the Web 2.0 platform for knowledge sharing and social interaction significantly influences nurses' decision to adopt it. On the other hand, since there may be some patient data privacy and confidentiality issues regarding the use of Web 2.0 tools for knowledge sharing [38], other issues important to nurses' decisions are hospital policy, regulations, and guidance on the use of Web 2.0 tools. Most importantly, the hospital always plays a leadership role in promoting and supporting nurses' adoption of new technology; thus, hospitals' leadership and support in constructing a Web 2.0 environment for knowledge sharing, learning, social interaction, and the production of collective intelligence are important to their decision. Therefore, the hospital's attitude is a major concern of nurses related to the adoption of Web 2.0 tools. In addition, senior management influence is also slightly relevant to nurses' decisions because

XSL•FO
RenderX

nurses require the support and encouragement of senior management to improve their nursing knowledge and learning. In summary, it can be concluded that peer participation and hospital support with policy and regulation on the use of Web 2.0 tools are the primary factors influencing their adoption by nurses and that senior management encouragement and support are secondary concerns.

The testing of the third set of hypotheses showed that the perceived behavioral control of human beings is positively correlated with resource (r = .41, P < .05) and technological conditions (r = .69, P < .05). The t value of the technology facilitating conditions (t = 18.78, P < .05) is higher than that of the resource facilitating conditions (t = 8.85, P < .05). This can be explained by the fact that nurses are mostly concerned about the availability of technology since Web 2.0 tools are new technology. Nurses are concerned about how and whether Web 2.0 functions can be accessed and used in Internet resources (eg, RSS feed reader) or their computing/mobile platform. This depends on the technology development of the Internet content or service providers or the technology infrastructure of the hospital environment [37]. By contrast, resource facilitating conditions such as money and time are less important to nurses when compared with technology facilitating conditions.

Testing the last set of hypotheses showed that usage behavior (r = .60, P < .05) is positively correlated with behavioral intention. Behavioral intention is positively correlated with attitude (r = .72,P < .05), subjective norm (r = .45,P < .05), and perceived behavioral control (r = .58, P < .05). The significance of the correlation between usage behavior and behavioral intention (t = 14.77, P < .05) is high. The significance of the constructs between behavioral intention and attitude (t = 20.20, P < .05), perceived behavioral control (t = 14.02, P < .05), and subjective norm (t = 9.81, P < .05) are in descending order. The result of testing the hypothesis regarding usage behavior is similar to the finding of Ajzen [26] that the three constructs are also correlated with behavioral intention. Thus, it can be concluded that the significant priorities of nurses' concerns regarding the adoption of Web 2.0 tools are attitude, perceived behavioral control, and subjective norm.

In conclusion, the primary concerns regarding the adoption of Web 2.0 tools are usefulness, advantages, compatibility, and technology availability and the secondary concerns are resource facilitating conditions and peer, hospital, and senior management attitude. The implication, then, is that health policy makers should make more effort to illustrate the usefulness, advantages, and compatibility of the application of Web 2.0 tools for knowledge sharing, learning, social interaction, and the production of collective intelligence and ensure that the technology is available to nurses. The other work for policy makers is to take a leadership role in promoting and supporting the adoption of Web 2.0 tools in the hospital environment and encouraging nurses to adopt Web 2.0 tools with their peers and senior management. Other resources such as money, time, and trainers can be subsidized or provided by the hospital authority.

Because physicians, nurses, and other health care professionals have close interaction, collaboration, and communication with each other on medical assessment, patient care and therapy, then knowledge sharing, learning, social interaction, and the production of collective intelligence are important for them to improve their skills and deliver a higher quality of medical service. Since Web 2.0 tools provide a platform to connect all these professionals together for knowledge sharing, learning, social interaction, and the production of collective intelligence, health policy makers can extend the use of Web 2.0 tools to these professionals. Therefore, using the DTPB can help them to identify their concerns regarding the adoption of Web 2.0 tools and to define strategies for promoting Web 2.0 in the whole hospital environment.

Acknowledgments

The author completed the majority of the work while she was serving the Hong Kong Polytechnic University and completed part of the work while at the the Hong Kong University of Science and Technology. Thus, the author would like to thank the Hong Kong Polytechnic University for the research environment and would also like to thank the 10 final year project students at the School of Nursing of the Hong Kong Polytechnic University for their assistance with the data collection. Thanks also go to the professors of the Nethersole School of Nursing at the Chinese University of Hong Kong for distributing and collecting the questionnaire in their programs.

Conflicts of Interest

None declared

Multimedia Appendix 1

Questionnaire for adoption of Web 2.0 tools for knowledge sharing, learning, social interaction, and collective intelligence.

[PDF File (Adobe PDF File), 36KB - jmir_v13i4e92_app1.pdf]

References

1. Santoro E. [Podcasts, wikis and blogs: the Web 2.0 tools for medical and health education]. Recenti Prog Med 2007 Oct;98(10):484-494. [Medline: 17970173]

XSL•FO
RenderX

2. Gooding J. Web 2.0: A vehicle for transforming education. International Journal of Information and Communication Technology Education 2008;4(2):44-53. [doi: 10.4018/jicte.2008040104]
3. Zejicek M. Web 2.0: hype or happiness? New York, NY: ACM; 2007 Presented at: 2007 International Cross-Disciplinary Conference on Web Accessibility (W4A 2007); May 7-8, 2007; Banff, Canada. [doi: 10.1145/1243441.1243453]
4. Mason R, Rennie F. Using Web 2.0 for learning in the community. The Internet and Higher Education 2007;10(3):196-203. [doi: 10.1016/j.iheduc.2007.06.003]
5. Ballantyne NJ, Quinn K. Front Page. San Mateo, CA: PBworks; 2006 Jun. Informal learning and the social web, part 1 URL: http://informallearning.pbworks.com/w/page/19914104/FrontPage [accessed 2011-09-22] [WebCite Cache ID 61sxGQui3]
6. Konstantinidis ST, Konstantinidis E, Nikolaidou MM, Boutou AK, Havouzis N, Argyropoulou P, et al. The use of open source and Web 2.0 in developing an integrated EHR and e-learning system for the Greek Smoking Cessation Network. Stud Health Technol Inform 2009;150:354-358. [Medline: 19745330]
7. Burrell AR, Elliott D, Hansen MM. ICT in the ICU: using Web 2.0 to enhance a community of practice for intensive care physicians. Crit Care Resusc 2009 Jun;11(2):155-159. [Medline: 19485881]
8. Lau A, Tsui E. Application of Web 2.0 technology for clinical training. In: Lytras MD, Ordóñez de Pablos P, editors. Social Web Evolution: Integrating Semantic Applications and Web 2.0 Technologies. Hershey, PA: Information Science Reference; 2009:132-137 URL: http://repository.lib.polyu.edu.hk/jspui/handle/10397/1452
9. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
10. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]
11. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23. [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
12. Boulos MN, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. BMC Med Educ 2006;6:41 [FREE Full text] [doi: 10.1186/1472-6920-6-41] [Medline: 16911779]
13. McGee JB, Begg M. What medical educators need to know about "Web 2.0". Medical Teacher 2008;30(2):164-169. [doi: 10.1080/01421590701881673]
14. McLean R, Richards BH, Wardman JI. The effect of Web 2.0 on the future of medical practice and education: Darwikinian evolution or folksonomic revolution? Med J Aust 2007 Aug 6;187(3):174-177 [FREE Full text] [Medline: 17680746]
15. O'Neill PA, Willis SC, Jones A. A model of how students link problem-based learning with clinical experience through "elaboration". Acad Med 2002 Jun;77(6):552-561. [Medline: 12063203]
16. Sandars J, Schroter S. Web 2.0 technologies for undergraduate and postgraduate medical education: an online survey. Postgrad Med J 2007 Dec;83(986):759-762. [doi: 10.1136/pgmj.2007.063123] [Medline: 18057175]
17. Connor E. Medical librarian 2.0. Med Ref Serv Q 2007;26(1):1-15. [doi: 10.1300/J115v26n01_01] [Medline: 17210545]
18. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
19. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
20. Pawlak R, Brown D, Meyer MK, Connell C, Yadrick K, Johnson JT, et al. Theory of planned behavior and multivitamin supplement use in Caucasian college females. J Prim Prev 2008 Jan;29(1):57-71. [doi: 10.1007/s10935-008-0127-y] [Medline: 18386181]
21. Kim D, Chang H. Key functional characteristics in designing and operating health information websites for user satisfaction: an application of the extended technology acceptance model. Int J Med Inform 2007;76(11-12):790-800. [doi: 10.1016/j.ijmedinf.2006.09.001] [Medline: 17049917]
22. Shin DH, Kim WY, Kim WY. Applying the Technology Acceptance Model and flow theory to Cyworld user behavior: implication of the Web 2.0 user acceptance. Cyberpsychol Behav 2008 Jun;11(3):378-382. [doi: 10.1089/cpb.2007.0117] [Medline: 18537514]
23. Davis FD. Perceived usefulness, perceived ease of use and user acceptance of information technology. MIS Quarterly 1989;13(3):319-340. [doi: 10.2307/249008]
24. Iacovou C, Izak B, Dexter A. Electronic data interchange and small organizations: Adoption and impact of technology. MIS Quarterly 1995 Dec;19(4):465-485. [doi: 10.2307/249629]
25. Benbasat I, Wang W. Trust in and adoption of online recommendation agents. Journal of the Association for Information Systems 2005;6(3):72-101. [doi: 10.1016/j.jsis.2007.12.002] [PMC: 16683729]
26. Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. Englewood Cliffs, NJ: Prentice-Hall; 1980.
27. Fishbein M, Ajzen I. Attitudes towards objects as predictors of single and multiple behavioral criteria. Psychological Review 1974;81(1):59-74. [doi: 10.1037/h0035872]

XSL•FO
RenderX

28. Leone L, Perugini M, Ercolani AP. A comparison of three models of attitude-behaviour relationships in the studying behaviour domain. European Journal of Social Psychology 1999;29(2-3):161-189. [doi: 10.1002/(SICI)1099-0992(199903/05)29:2/33.0.CO;2-G]
29. Davis FD. A Technology Acceptance Model for Empirically Testing New End-user Information Systems: Theory and Results [dissertation]. Cambridge, MA: Massachusetts Institute of Technology; 1986. URL: http://dspace.mit.edu/handle/1721.1/15192 [accessed 2011-09-24] [WebCite Cache ID 61wJnDUY2]
30. Lederer AL, Maupin DJ, Sena MP, Zhuang Y. The technology acceptance model and the World Wide Web. Decision Support Systems 2000;29(3):269-282 [FREE Full text] [doi: 10.1016/S0167-9236(00)00076-2]
31. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Management Science 1989;35(8):982-1003 [FREE Full text] [doi: 10.1287/mnsc.35.8.982]
32. Taylor S, Todd PA. Understanding information technology usage: a test of competing models. Information Systems Research 1995;6(2):144-176 [FREE Full text] [doi: 10.1287/isre.6.2.144]
33. Shih Y, Fang K. The use of a decomposed theory of planned behavior to study Internet banking in Taiwan. Internet Research 2004;14(3):213-223. [doi: 10.1504/EG.2007.012177]
34. Lau ASM, Kwok VWS. How e-government strategies influence e-commerce adoption by SMEs. Electronic Government, an International Journal 2007;4(1):20-42.
35. Fink A. How To Analyze Survey Data. Thousand Oaks, CA: Sage Publications; 1995.
36. Dillon TW, McDowell D, Salimian F, Conklin D. Perceived ease of use and usefulness of bedside-computer systems. Comput Nurs 1998;16(3):151-156. [Medline: 9611866]
37. Leonard, 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 [FREE Full text] [doi: 10.2196/jmir.9.2.e9] [Medline: 17513287]
38. Hughes B, Joshi I, Wareham J. Health 2.0 and Medicine 2.0: tensions and controversies in the field. J Med Internet Res 2008;10(3):e23 [FREE Full text] [doi: 10.2196/jmir.1056] [Medline: 18682374]

Abbreviations

DTPB: decomposed theory of planned behavior
IP: Internet protocol
IT: information technology
RSS: really simple syndication
TAM: technology acceptance model
TPB: theory of planned behavior
TRA: theory of reasoned action
UB: usage behavior

Edited by G Eysenbach; submitted 05.11.09; peer-reviewed by S Erdley; comments to author 10.02.10; revised version received 06.05.10; accepted 11.11.10; published 11.11.11.

Please cite as:
Lau AS
Hospital-Based Nurses' Perceptions of the Adoption of Web 2.0 Tools for Knowledge Sharing, Learning, Social Interaction and the Production of Collective Intelligence
J Med Internet Res 2011;13(4):e92
URL: http://www.jmir.org/2011/4/e92/
doi:10.2196/jmir.1398
PMID:22079851

©Adela S.M. Lau. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 11.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Online Social Networks and Smoking Cessation: A Scientific Research Agenda

Nathan K Cobb[1,2,3,4], MD; Amanda L Graham[1,3], PhD; M. Justin Byron[4], MHS; Raymond S Niaura[1,3,4], PhD; David B Abrams[1,3,4], PhD; Workshop Participants[5]

[1]The Schroeder Institute for Tobacco Research and Policy Studies, American Legacy Foundation, Washington, DC, United States
[2]Division of Pulmonary & Critical Care, Department of Medicine, Georgetown University Medical Center, Washington, DC, United States
[3]Department of Oncology, Georgetown University Medical Center / Lombardi Comprehensive Cancer Center, Washington, DC, United States
[4]Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
[5]see acknowledgements

Corresponding Author:
Nathan K Cobb, MD
The Schroeder Institute for Tobacco Research and Policy Studies
American Legacy Foundation
1724 Massachusetts Ave NW
Washington, DC, 20036
United States
Phone: 1 202 454 5745
Fax: 1 202 454 5785
Email: ncobb@legacyforhealth.org

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2012/1/e12

Abstract

Background: Smoking remains one of the most pressing public health problems in the United States and internationally. The concurrent evolution of the Internet, social network science, and online communities offers a potential target for high-yield interventions capable of shifting population-level smoking rates and substantially improving public health.

Objective: Our objective was to convene leading practitioners in relevant disciplines to develop the core of a strategic research agenda on online social networks and their use for smoking cessation, with implications for other health behaviors.

Methods: We conducted a 100-person, 2-day, multidisciplinary workshop in Washington, DC, USA. Participants worked in small groups to formulate research questions that could move the field forward. Discussions and resulting questions were synthesized by the workshop planning committee.

Results: We considered 34 questions in four categories (*advancing theory*, *understanding fundamental mechanisms*, *intervention approaches*, and *evaluation*) to be the most pressing.

Conclusions: Online social networks might facilitate smoking cessation in several ways. Identifying new theories, translating these into functional interventions, and evaluating the results will require a concerted transdisciplinary effort. This report presents a series of research questions to assist researchers, developers, and funders in the process of efficiently moving this field forward.

(J Med Internet Res 2011;13(4):e119) doi:10.2196/jmir.1911

KEYWORDS

Smoking cessation; social support; social networks; addiction; treatment; tobacco

XSL•FO
RenderX

Introduction

Smoking remains the leading cause of 443,000 preventable deaths and nearly US $200 billion in excess costs in the United States each year [1]. Smoking rates in the United States have stalled near 20%, [2] and large-scale reduction in smoking prevalence remains an urgent public health imperative. Although the evidence-based cessation interventions recommended by the US clinical practice guideline for tobacco dependence treatment [3] have been shown to double quit rates, they are largely underused [4]. Reaching the US public health goals of cutting the smoking rate to no higher than 12% by 2020 [5] will require novel approaches to create new interventions, enhance the effectiveness of existing cessation treatments, and maximize the reach and utilization of both.

The evolution of the Internet and the growth of online social networks may present a solution to the intertwined problems of effectiveness and reach of cessation interventions. Social support [6], social integration [7], and social networks [8] appear to play important roles in smoking behavior and cessation. Yet numerous tobacco treatment studies aimed at creating supportive relationships (eg, peer or buddy training) or harnessing existing social relationships (eg, spouse interventions) have generally yielded disappointing results [9-12]. The limitations of traditional treatment settings in which this work was conducted (eg, low attendance, time constraints, or type and number of available support persons) may partially explain the difficulty in leveraging social support in the cessation process.

Online social networks, by contrast, offer round-the-clock access to vast numbers of participants, potentially superseding these limitations and offering a realistic delivery model for social support. In theory, smokers might benefit not only from active, personal interactions with other network members, but also from various passive sources of social support and influence. Such interactions could alter an individual's motivation to quit, reinforce the undesirability of smoking, assist in buffering cessation-related stressors and enhancing coping skills, and provide suggestions for eliminating smoking cues [13]. To date, there has been a wealth of behavioral science research on the role of social networks in face-to-face interactions but little published research on online social networks [14-16].

The growth of online social networks and their penetration into popular awareness has been phenomenal, with over 70% of American adults now using some form of social media or online social network [17]. As of early 2011, an estimated 150 million Americans actively use Facebook, the largest of the online social networks [18]. Intentionally created online networks dedicated to smoking cessation are smaller but have been in existence for over a decade. These types of dedicated systems—where smokers and former smokers communicate through various channels in an effort to quit and stay abstinent—are now widely used by hundreds of thousands of smokers over relatively long periods of time [16,19]. Over the years, cessation-focused online networks have evolved from simple systems for the exchange of messages to complex networks complete with multiple modes of communication (eg, chat rooms, forums, or private messaging), self-representation (eg, personal profiles, blogs, or journals), and affiliations (eg, buddy or friends lists, or private groups), through which social norms, social influence, and social support may be conveyed in real time [16,20].

Concurrent with the exponential growth of online social networks has been the rapid evolution of social network science, spurred on as improvements in computer capacity and software have caught up with theory and the burgeoning size of available data sets [21,22]. In studies of real-world networks, social network science has demonstrated that social influence flows through networks and can influence a broad range of behavioral and emotional changes, including smoking and alcohol use [8,16,23], obesity [24,25], happiness [26], and depression [27], as well as loneliness [28] and suicide [29]. Social network analysis allows for an expanded view of an individual's social universe, taking into account not only their own connections, but also the connections of their friends and contacts and beyond. This ability to look at the social structure in aggregate allows for inferences about how topology (the network structure) both enables and drives behavior change.

Actually cutting smoking prevalence by nearly half by 2020 will require cessation interventions that can reach millions of people in consumer-friendly ways. The convergence of robust evidence for the role of social support in cessation, the growth and proliferation of online networks, and the recent advances in social network analytic techniques present an opportunity for the development and dissemination of high-impact interventions targeting smoking. The notion that online social networks present a powerful and novel approach to cessation is supported by a research in relatively disparate disciplines, including tobacco control, social psychology, and social network science, to name just a few. Leveraging the enormous potential of online social networks to reach and treat smokers will require a transdisciplinary conversation among researchers, developers, and funders that bridges behavioral, network, and computer sciences and other fields [30,31]. We sought to initiate this discussion by convening a multidisciplinary group of experts to identify gaps in knowledge and research questions regarding the potential of online social networks to more rapidly reduce smoking prevalence. Our goal was to construct a strategic research agenda to guide future collaborative work. This paper presents this agenda in the form of 34 pressing research questions and related issues, along with brief discussion.

Methods

We invited approximately 100 experts and thought leaders (listed under Acknowledgements at the end of this article) across a range of relevant content areas to a 2-day workshop held September 30 to October 1, 2010 in Washington, DC. Participants represented a broad range of disciplines, including economics, engineering, epidemiology, linguistics, mathematics, medicine, nursing, psychology, public health, network science, sociology, software engineering, and product design and commercialization. A small number of participants were invited to give focused overview presentations to help bridge disciplinary borders and to establish a common starting point for discussion. These included presentations on the epidemiology and treatment of tobacco use; basic principles of

XSL•FO
RenderX

social support theory and social support interventions in tobacco control; social network science and network-based interventions in tobacco control; the history, evolution, and current state-of-the-science of general and cessation-specific online social networks; and methodological, measurement, and analytic issues regarding social network data collection, analysis, and interpretation. Additionally, representatives from three of the largest for-profit, health-related, online social network interventions were invited to describe their programs and the lessons they had learned in managing online networks.

Following the overview presentations, participants were divided into small multidisciplinary working groups and tasked with developing a list of priority research questions. The guiding framework for workgroup discussions was to address the key question "What do we know and what do we need to learn that will make a difference in improving cessation outcomes?" The framing of the question was deliberately broad to enable participants from diverse disciplines and with varying content expertise to contribute their perspectives.

Participants were instructed to formulate and group research questions into four major categories: (1) *advancing theory* (developing, refining, or integrating existing theories and models from online and offline social network, social support, smoking cessation, and behavior-change domains), (2) *understanding fundamental mechanisms* (how online social networks produce behavior change at the individual and network level), (3) *intervention aproaches valuation* (methods and metrics for appropriate program, process, and outcome evaluations). To encourage brainstorming, groups were instructed to imagine they had access to the intellectual, financial, and technical resources represented by any combination of the attendees or speakers at the workshop. Each working group presented their list of research questions back to the full group for further discussion. The groups worked independently on each major theme area, and then their recommendations were synthesized and refined with feedback from the entire group. After the workshop, the planning committee met to review the findings and to summarize the general areas of research topics and priorities for dissemination.

Results

Participants generated a large number of research questions at varying levels of granularity. Common and overlapping ideas were integrated and a subset of questions was selected for further discussion and elaboration by the report's authors. For each key topic area, we present a summary of discussions and provide examples of the most pressing research questions or issues raised.

Several overarching themes emerged from the discussions. First, participants noted that traditional models of offline (eg, face-to-face) intervention and evaluation are often reflexively applied to online observations or interventions. While there are ways in which offline and online behaviors overlap and can reciprocally inform models, mechanisms, implementation, and evaluation, there are also important differences that require critical thinking about online networks. There is a need to challenge and test the assumptions inherent in traditional models when developing, implementing, and evaluating online interventions.

A second theme related to the mechanisms of behavior change. Numerous theory-based processes of behavior change have been described within social networks, including diffusion of information, viral spread of interventions, social support, social norms, and modeling. It is unknown whether these or other unidentified processes are important in online social networks for cessation, and if any of these may be iatrogenic (ie, promoting continued smoking rather than cessation).

A third theme centered on the appropriate use of theoretical models, empiricism, and statistical or simulation modeling techniques. Future advances in online social network interventions will likely depend on a transdisciplinary approach to develop appropriate theoretical models, test them in vivo and in silico (software modeling), rapidly iterate to determine interventions with the highest probability of effect, and perform intervention trials with appropriate research designs and end points. Such advances will require improvements in existing capacity to collect complex and large-scale longitudinal data on behaviors and interactions within online networks.

Finally, we note a common assumption during the workshop that social network interventions will increasingly take advantage of mobile delivery mechanisms—whether smart phone apps, text messages, or other formats. While few questions address this shift explicitly, we have attempted to write this summary to be agnostic toward delivery platform. Both questions and recommendations are intended to be broadly applicable, regardless of location or modality.

Advancing Theory of How Social Networks Influence Smoking-Cessation Behavior

Social network and social support models in smoking cessation [6] derive primarily from social learning theory [32] and from the study of interventions to change existing social support interactions or develop new supports (eg, from a counselor and other members of a group participating in face-to-face smoking-cessation treatments) [10,33]. Observational studies have shown that social support is associated with smoking initiation and cessation, and that smokers associate with other smokers in proximal social networks. Intervention studies, though, have yielded mixed and largely disappointing results in attempting to manipulate or harness support [6,10,33]. These findings highlight the importance of the need for a theoretical framework that permits simultaneous understanding of the observed association between social phenomena and smoking, and manipulation of the social environment to effect change. There is a need to evolve our theories and models to refine their explanatory power and their applied utility to facilitate behavior changes, such as smoking cessation [34].

The application of network theory to social networks has largely occurred in studies of real-world (ie, offline) networks [8,29,35-37] using retrospective self-report measures and cross-sectional data. Recent research on social networks has rapidly evolved using online data [16,38-42], new computational methods, and mathematical and simulation modeling [37,43,44]. Data from electronic communications networks (eg, online

XSL•FO
RenderX

social networks, email systems, and telephone networks) can be collected in real time and can record communications and interactions at multiple levels and with repeated observations of intraindividual, interindividual, and contextual influences. Such methods can inform interventions and measures of process and outcome, but the proliferation of data and results calls for new or more refined models or theoretical frameworks to facilitate interpretation and application.

Theories that try to explain and change behavior in small real-world settings may not translate easily into the online world, where interactions occur on a larger scale and in a different medium. A transdisciplinary synthesis [30,45] is likely required to integrate our understanding of the nature and form of social networks, as instantiated or reflected in the online world, and their functions that serve to initiate and maintain changes in individual behaviors. Structuralist social network theories, which address how patterns of social relationships are associated with substantive topics such as health behaviors [37,45], come closest to fusing form and function, and serve as a useful point of departure for understanding how social networks and individual tobacco use behaviors intersect.

A century ago, Simmel [35] called for more than knowing how to measure characteristics of networks, such as the density of their interconnections, and recommended developing a set of assumptions about how best to describe and explain the social phenomenon of interest in its proper context. This challenge remains today as we seek to integrate social and individual theories of behavior change. Indeed, the structure of a network may induce, maintain, or strengthen a behavior not just by transmission of information, but via forces of exclusion, adaptation, or the binding together of members [46]. Looking at network causes of phenomena of interest requires asking what kinds of social networks lead to particular outcomes.

Specific Questions

1. How well do theoretical models of social influence translate between offline and online contexts?

2. How does online social network data map onto real-world networks? Does research based on retrospective self-report with sparse observations in the real-world match with dynamic, observed behavioral data collected online?

3. How can behaviorally important ties be identified in online social networks that may be composed of large numbers of apparently weak ties?

Understanding Fundamental Mechanisms

Online social networks exist across a broad range of health conditions and behavioral risk factors including tobacco use (eg, becomeanex.org, quitnet.com, and stopsmokingcenter.com), diet and fitness (sparkpeople.com), diabetes (tudiabetes.com), chronic diseases (patientslikeme.com), and others [16,47-49]. Research is in its infancy regarding the mechanisms through which these online social networks might or might not effect behavior change. Social support models suggest that behavior change is mediated in part through information exchange, instrumental or emotional support, stress buffering, or improved self-efficacy [6]. However, other mechanisms may be as or more important in online networks, such as exposure to new or different norms or behaviors modeled by other network members [16]. To date, the design and implementation of online networks has been largely based on offline cessation approaches, usually comprising only small groups of smokers actively trying to quit. The evolution of more effective cessation interventions will require an in-depth, sophisticated understanding of the unique aspects of online social networks and the specific mechanisms through which they effect behavior change, as well as the careful selection of evaluation strategies matched to this intervention context.

Homophily, Heterophily, and Network Topology

Homophily refers to the tendency of people to associate with similar others (“birds of a feather”), while heterophily refers to the tendency to collect in diverse groups. That homophily tends to be a driving factor in the formation of social networks [50] is an important consideration in offline networks: the tendency of smokers to associate with other smokers may decrease the impact of normative exposure to nonsmokers or former smokers within a network. In contrast, online social networks may be heterophilous, comprising individuals across the cessation continuum including individuals who have been abstinent for years [16] or current smokers who are curious but not yet motivated to quit. Research in offline networks suggests that topological factors (the pattern of ties between individuals within the network), such as clustering of smokers, affects cessation over time [8]. Other work in online networks indicates that dense connections at the individual level reinforce social signaling and increase the chance of behavior change [39]. As most existing online networks remain uncharacterized, little is known about their structure or the optimal topology to effect behavior change.

Specific Questions

1. What is the role of homophily in the formation of online networks?

2. What is the role of heterophily in the provision of social support throughout the cessation process, and how does it influence cessation outcomes?

3. Can ties within online social networks be fostered or manipulated to “rewire” networks, modify topology, and drive behavior change?

4. What impact does network topology have on behavior change? For example, does having a dense local network increase the probability of making a quit attempt, cessation, and maintenance of abstinence?

Social Diffusion

Information and behavior diffusion through offline social networks are well-studied phenomena, encompassing myriad behaviors from seed choice by farmers to the spread of smoke-free policies from city to city [51]. In contrast, inducing or manipulating diffusion through both online [40,41] and offline networks [51,52] has proved challenging in practice; deliberately causing spread of information or a behavior is easier to conceptualize than to implement. In commerce and industry, the term viral marketing refers to this deliberate seeding and

XSL•FO
RenderX

resulting diffusion of a message through a targeted network, such as the promotion of a new product [53]. While viral marketing is a common practice, there is little academic literature on its use for health topics or for online approaches for health behavior change. Nonetheless, deliberate seeding and diffusion may allow for the dissemination through created networks of specific information (eg, information about a new cessation medication), interventions (eg, a quit smoking app through Facebook), smoking/cessation norms, and other health behaviors.

Specific Questions

1. How does information spread through an online social network? Are there identifiable patterns of information spread that can be leveraged in intervention research?

2. Can key participants in a network be identified and targeted to foster information diffusion or make it more efficient?

3. What are the drivers of the viral spread of an application, concept, or innovation through online networks?

4. How does network topology affect diffusion? Can social network measures and concepts such as centrality or clustering be used to predict or alter diffusion?

Social Norms and Modeling Behavior

Despite the fact that members do not know each other at the outset, created online networks can develop their own language and norms [54]. Existing members may convey expectations for certain behaviors or participation in the network that guide and support new or struggling members [55]. These expectations and norms may differ from those in the participant's offline network. For example, the public health community has worked hard to normalize the use of nicotine replacement as a cessation aid; however, most smokers do not use pharmacotherapy when quitting [4] and many have concerns about the safety of any form of nicotine [56]. Online social networks may present norms supportive of medication use, and existing users may model successful medication use behavior. Other norms such as recycling after failed quit attempts, enlisting external social support, or the use of telephone counseling are other examples of potential norms (positive or negative) within social networks.

Specific Questions

1. How are social norms established and communicated in an online social network?

2. What is the effect of online social norms when they differ from a user's offline environment?

3. Does anonymity in online networks enhance or diminish the effect of modeling behavior and communication of norms?

4. Are norms and modeling effective mechanisms to influence "lurkers" (ie, members of a network that read other members' posts/comments but rarely communicate with other members)?

Network Formation, Social Integration, Retention, and Longitudinal Stability

There are numerous online communities and created social networks dedicated to health-related behavior change—some of them in existence for over a decade with thousands of members—yet it remains unclear what factors led to their growth or stability. Previous research has shown that small numbers of individuals may be responsible for approaching and "integrating" new members as they join an online network for cessation [16]. Most research has reported results from successful, stable networks [14,16,57], while projects that fail to form networks are rarely reported [58]. As a result, the factors that drive member integration and retention and network stability remain unclear. Adequate understanding of these factors is required to build new interventions and to maintain existing versions or enhance their effectiveness.

Specific Questions

1. What predicts engagement in an online social network? What demographic, smoking, psychosocial, or other characteristics are predictive of participation and integration?

2. What is the role of timing of interactions in online social network in influencing integration and participation? What forms of outreach and communications (eg, private messages, instant messaging, public forums, or blogs) drive tie formation?

3. What is the role of long-term users in network structure and network stability over time?

Intervention Design and Approaches

The incredible growth of online social networks offers the opportunity for novel intervention designs. Created networks such as online communities dedicated to smoking cessation are a common component of modern health behavior-change systems and often center on the "build it and they will come" premise of intervention delivery. These networks generally comprise motivated individuals ready to make or maintain changes to one or more health risk behaviors. Such systems benefit from a specific focus, on the part of both the user and intervention designers. However, they generally do not yet take full advantage of the potential to proactively reach larger populations. Individuals must generally seek out and enroll in the closed system, and ultimately many registrants fail to return to the site [59], much less engage with the social aspects as designed (they become, at best, lurkers or, at worst, completely unengaged). In contrast, general-purpose networks such as Facebook offer unique opportunities and challenges, related primarily to their enormous size, including the potential for autonomous propagation (viral spread) of interventions. Certainly, intervention design decisions should be informed by relevant and sophisticated theories that specify the active ingredients and mechanisms of action, but the surfeit of potential participants in these extremely large networks ultimately allows for data-driven methods to drive the ongoing design and refinement of interventions.

Target Populations

Smoking-cessation interventions most frequently target individuals ready to make a quit attempt. Yet many people who join online cessation systems have already quit or are not ready to make a quit attempt [60-63]. Traditional social support models will need to be modified to assist these individuals and to maximize their utility in supporting others. Significant public and private resources are used each year to denormalize smoking and encourage cessation using traditional media [64,65]. As

XSL•FO
RenderX

public health organizations increasingly use advertising and outreach efforts to drive utilization of online resources, it will become imperative to identify the types of smokers that may benefit from social network-based approaches to cessation. A one-size-fits-all approach is unlikely to be efficient or effective, and it is unclear how much customization or individual tailoring is needed to make an incremental addition to outcomes [66,67].

Specific Questions

1. Do smokers who are not motivated for behavior change benefit from social network interventions? What influence do social support and normative exposures have on smokers who may not be thinking of quitting?

2. Can online social networks assist smokers who have already quit to maintain abstinence? Can recruiting abstinent smokers into a network strengthen the network's capacity for social support?

3. Are demographic or psychosocial characteristics important predictors of online social network utilization? What is the impact of age, gender, race/ethnicity, or other identifying characteristics with regard to network phenomena such as integration or tie formation?

4. How can network-based interventions capitalize on secular trends and historical events, such as a change in the federal excise tax rate, new year's resolutions, The Great American Smokeout, or major smoking-related media stories such as the death of Peter Jennings from lung cancer? Do smokers recruited during the "surges" associated with these events differ from those who join an online social network at other times or for other reasons?

Systems Integration

The oldest examples of online social networks for cessation are relatively siloed intervention approaches, focused largely on engaging users with other participants on a cessation-specific website and in an anonymous fashion. More recent interventions integrate online social networks into other treatment-delivery approaches, such as telephone quitlines [13,68,69]. The rapid expansion of large-scale networks such as Facebook where users are personally identifiable offers the opportunity to disseminate cessation interventions through existing networks, but without the aspect of anonymous participation. It is unknown the degree to which the advantages of leveraging an existing network where participants are identified are offset by the potential benefits of a network where members are anonymous. Integration of online social networks with other treatment modalities (eg, text messaging, health care-delivery settings, or electronic medical record systems) offers the opportunity to enhance treatment effectiveness, augment social support mechanics, and increase the reach of traditional services. At the same time, such integration introduces multiple new complexities.

Specific Questions

1. What is the best mechanism for online social networks to interface with other elements of health care or tobacco treatment (eg, telephone quitlines, over-the-counter and prescription pharmacotherapy, physician advice, electronic medical record, mass media campaigns, or policies)?

2. How does involving a smoker's offline network (eg, friends, family, medical practitioners, worksite wellness, or occupational health programs) augment or diminish the effect of an online social network on cessation?

Development Methods

There is a chasm between the rapid-cycle, diffusion-focused development methods used by entrepreneurs and industry to launch online programs and the traditional, efficacy-based development methods of behavioral and social scientists. For example, Facebook has grown literally from a dorm room project to over 150 million Americans a month in approximately 6 years. Ironically, this is typically the same amount of time between submission of a federal grant application and the publication of its main outcome paper. Shortening this timeline is critical if we are to develop effective interventions that can be deployed on a large scale to benefit public health in a timely fashion. Engineering principles of iterative development and early evaluation have been adapted in the behavioral sciences (eg, multiphase optimization strategy, or "MOST", [70]) and provide one approach to achieve this goal. Online interventions are particularly suited to these methods; large, available target populations enable intervention variations to be tested against each other with statistical significance in rapid sequence or in factorial models, in theory improving effectiveness and tightening research and development timelines [67,70].

Specific Questions

1. Can engineering models, such as MOST, speed development time and/or increase efficacy of network-based interventions?

2. What process and outcome metrics are most appropriate during intervention development and refinement? Participant engagement? Retention? Network integration? Quit attempts? Early abstinence?

Evaluation

Several high-quality randomized controlled trials of Internet cessation programs have been conducted [13,57,62,67,71-76]. Yet, to date, there have been no published reports of tobacco intervention trials that link social network structure or dynamics to either social support metrics or more distally to cessation outcomes. Not all of these trials have included social network components, but among the ones that did, there are several reasons for this gap in the literature: the difficulty of constructing appropriate assessment and intervention protocols, the difficulty in maintaining participants in social interventions, and the challenge in disentangling social processes from other features of many Web-based interventions (eg, tailored materials, expert systems, or access to counseling staff). There is a critical need for the identification of appropriate research designs, data collection methods, and evaluation strategies to determine the impact of social processes within online interventions that may drive cessation and abstinence.

Research Design

The use of randomized control trials in research to evaluate online social network-based interventions presents a number of challenges. Among these are selecting a feasible, ethical, and rigorous control condition [77,78]; avoiding contact between

XSL•FO
RenderX

participants randomly assigned to different conditions; and managing the attrition observed across virtually all online interventions [59]. Alternative evaluation designs used in eHealth research, such as practical clinical trials, pragmatic randomized controlled trials, and nonexperimental and quasi-experimental designs [77,79-81], may be appropriate as well for social network interventions. Given the size of data sets that are generated from online social network interventions, automated systems for the categorization and extraction of data (such as natural language processing and sentiment analysis, data mining, and pattern recognition) may also play important roles in exploratory analyses. The use of varied methods and data sets will make consistent and standardized reporting of results increasingly important as the field advances.

Specific Questions

1. Given that alternative Internet interventions are a mouse-click away, what are the important considerations in selecting a rigorous and appropriate control group and evaluating contamination (ie, exposure to the intervention arm among control participants)?

2. Other than randomized control trials, what rigorous research designs can be aptly used to optimize online social network interventions? Are there specific research designs that are best used at specific phases of the development–dissemination–implementation continuum?

Data Collection and Analysis

Online interventions and social networks in particular are part of the "big data" problem [22], an emerging issue where the quantity of behavioral and other process data exceeds the capacity for traditional analysis. Academic computational social science—the collection and analysis of these data—lags behind other fields such as physics and biology, as well as the corporate capacity of Google and Facebook in managing big data [21]. The two primary challenges inherent in big data are adequately defining and capturing the appropriate data, and conducting effective and efficient analyses. Data collection methods such as ecological momentary assessment, mobile tracking data, content and sentiment analysis, and observation of online interactions can provide granular information about behavior with minimal impact on the user or their friends and contacts. These methods can generate much richer—and also more complicated—representations of social networks that contain information about the weight of ties, their valence (positive or negative), and the presence of hidden or latent ties [82].

Specific Questions

1. How can novel data collection methods such as ecological momentary assessment, passive tracking data from websites, or data from mobile devices be used to gather network-level data without affecting individual behavior or the network itself?

2. What new techniques and analytic methods will be required for analysis of "big data" and increasingly complicated network representations?

Expanded Outcomes and End Points

Traditionally, research has evaluated the impact of an intervention only on the individuals enrolled in a study. Bolstered by evidence from both offline [8,23-29,51,52] and online [39-42,53] studies, network theory suggests that behavior change may diffuse through a network. Successful intervention with an individual smoker may have positive externalities (a term for collateral effects, drawn from the economic literature) that ripple through the network [83] causing other smokers to quit or to cut back on their smoking, or resulting in changes in attitudes or other beliefs [8]. For example, a quit attempt by an individual enrolled in a program might prompt a close friend to also attempt cessation. Success of the friend would not normally be part of a traditional analysis, but becomes critical from a network standpoint, particularly since interventions may be specifically designed to elicit this effect. Given that evaluating changes in behavior among individuals outside the purview of a research study may be difficult or impossible, alternative end points, outcomes, and evaluation strategies become imperative [83].

Specific Questions

1. What end points or surrogate outcomes will permit the evaluation of externalities in online network interventions?

2. What are the ethical implications of observing or even inducing behavior change in individuals that have not consented to participate in a research study?

Modeling to Inform Design and Evaluation

The use of mathematical predictive models in public health, and tobacco control in particular, has recent support [84-86]. Their use to design, refine, or evaluate behavioral interventions for cessation is less defined, but the opportunities are compelling. Previously, models have been employed to examine how best to optimize the multiple modes of delivery of smoking-cessation interventions, as well as to capitalize on context, such as multilevel influences of restrictive polices, mass media, and increased sales taxes [85,87-89]. In silico techniques such as agent-based modeling, where powerful computers simulate autonomous users interacting within a network over time, can be used to predict responses to intervention design changes [90]. Under certain circumstances they can also be used to disentangle behavioral outcomes from network processes and potentially contribute to evaluation [34]. Such techniques not only may play a valuable role in accelerating intervention development and evaluation, but also may help to determine the potential impact of interventions prior to time consuming and costly promotion and implementation–dissemination efforts.

Specific Questions

1. How can mathematical and computer-driven simulations of various kinds (eg, dynamic systems models or agent-based models) contribute to intervention development, refinement, or evaluation?

2. How can existing systems models inform work with online social networks? How might existing systems models be affected or informed by large-scale social networks (such as Facebook)?

XSL•FO
RenderX

Discussion

An increasingly interconnected online social Web provides incredible opportunities to shift behavior, affect health, and meet public health challenges. Despite promising starts in individual fields, it will take further rigorous and transdisciplinary research and development to meet the potential described in this report. Tackling the questions posed here, structuring research protocols, and developing appropriate analytic techniques will require true collaboration across multiple fields and divergent disciplines [30]. Success may lead to interventions with the capacity to reach large populations, augment existing treatment modalities, and effect behavioral change in novel ways.

While we have focused on tobacco use and smoking cessation, the same questions and approaches may apply to virtually any behavior change of interest. Interventions need to be informed by and should inform theory, model testing, and protocols for refinement. As we gain experience working in transdisciplinary teams and refine our models, we will have a clearer picture of the new measures needed for empirical data collection and testing of models to identify the mechanisms, pathways, and key processes that influence intermediate and final behavior-change outcomes of interest. Such iterative approaches will also lead to ways to validate self-report measures and integrate or triangulate the tracking of online activities with observational data and social network and support activities that are conducted offline.

Given the rapid evolution of the field of online communications and smoking-cessation interventions, and the numerous disciplines involved, we will need more agreement and standardization on metrics. For example, assessing norms and answering questions about their impact on behavior will require the development and validation of new instruments to determine active norms in an online social network and their importance. This work will be a necessary precursor to any efforts to modify existing norms or introduce new norms into existing or evolving networks. Ultimately the refinement of theories, models, and interventions would benefit from the development of standardized measures not only for norms, but for virtually all metrics mentioned in this report. Such measures would ideally have good reliability and validity across different projects, organizations, and even disciplines. Establishing a set of core measures that should be used across studies of online social networks will help test and improve both internal and external validity and will enhance theory testing by ensuring robustness, generalizability, replicability, consistency, and convergent validity across studies.

There are several limitations to this report. The recommendations presented are dependent on the individuals present at the conference and the structure provided by the organizers. Different participants or a different structure undoubtedly would have produced different questions and topics. The research priorities and recommendations presented here are but one set of views that we hope will serve to stimulate additional dialogue and research efforts. Addressing the questions posed in this report will present significant, but not insurmountable, challenges around personal privacy and the ethical treatment of research participants and their social contacts. Behavioral and biomedical researchers have traditionally thought about the impact on individuals, but social network interventions will challenge us to draw on the experience of public health professionals, social marketers, and sociologists as we increasingly target networks.

Networks and technology evolve on their own timeline, independent of the needs, funding, or aims of researchers. The study of rapidly evolving networks will require investigators and funders to tighten their timelines through the entire process (from idea, to funding, to execution, to publication). The traditional models of funding research via federal grants such as those in place at the National Science Foundation or National Institutes of Health in the United States are notoriously slow compared with industry and entrepreneurial interests. Network science and online interventions are changing rapidly and the traditional funding models must adapt as well. In 2009, Lazer and colleagues voiced concerns that research on large-scale networks “could become the exclusive domain of private companies and government agencies” [21], an outcome they noted would not be in the public interest. Developing and maintaining a strong academic research program is imperative and will require adjustments by funders, researchers, and publishers of scientific research.

Research efforts designed to address the topics and questions in this report may help identify mechanisms to significantly decrease the burden of tobacco related disease in the United States and elsewhere. The core ideas and themes developed here for smoking cessation may also apply—recognizing differences in context—to a variety of behaviors (eg, obesity, substance abuse, or adherence to medical recommendations) that could directly or indirectly improve the well-being and quality of life of our society. It is important to recognize that the powerful forces and rapid transmission of information across networks may also be used inappropriately or destructively (both intentionally and unintentionally) as well as for doing good. Ultimately, we hope that the kinds of research efforts encouraged in this paper will give rise to a new generation of interventions to help people quit smoking and stay quit, delivered and spread through a variety of social networks—networks that we recognize today, and networks that will develop tomorrow.

Acknowledgments

The workshop was funded by Healthways Inc., the National Cancer Institute of the National Institutes of Health, and the Johns Hopkins Bloomberg School of Public Health in conjunction with the American Legacy Foundation. External sponsors had no role in workshop planning or in the writing or publication of this report. The authors wish to thank all of the participants for their

XSL•FO
RenderX

time and effort and extremely valuable contributions. We especially appreciate the speakers and planning committee without whom the workshop would not have been possible.

List of workshop participants:

David Abrams*, PhD; Schroeder Institute for Tobacco Research & Policy Studies, Legacy

Jas Ahluwalia, MD; University of Minnesota

Larry An, MD; University of Michigan

Eric Asche; Legacy

Audie Atienza, PhD; National Institutes of Health

Erik Augustson*, PhD, MPH; National Cancer Institute, NIH

Cathy Backinger*, PhD, MPH; National Cancer Institute, NIH

Cathy Baker, PhD, RN; Case Western Reserve University

Carla Berg, PhD; Emory University

Greg Bloss, MA; National Institute on Alcohol Abuse and Alcoholism, NIH

Georgiy Bobashev, PhD; RTI International

Janet Brigham, PhD; SRI International

Joanne Brown, ARNP; University of Kentucky

Taneisha Buchanan, PhD; University of Minnesota

David Buller, PhD; Klein Buendel, Inc.

M. Justin Byron*, MHS; Schroeder Institute for Tobacco Research & Policy Studies, Legacy

Chris Cartter; MeYou Health, LLC

Damon Centola, PhD; Massachusetts Institute of Technology

Nicholas Christakis, MD, MPH, PhD; Harvard Medical School

Nathan Cobb*, MD; Schroeder Institute for Tobacco Research & Policy Studies, Legacy

Sheldon Cohen*, PhD; Carnegie Mellon University

Trevor Cohen, MBChB, PhD; The University of Texas

Linda Collins, PhD; Penn State University

Noshir Contractor*, PhD; Northwestern University

Mary E. Cooley, PhD, RN; Dana Farber Cancer Institute

Laurel Curry; Legacy

Lowell C. Dale, MD; Mayo Clinic Tobacco Quitline

RaeAnne Davis, MSPH; North American Quitline Consortium

Nathan Eagle, PhD; The MIT Design Laboratory, Massachusetts Institute of Technology

Jason Fletcher, PhD; Yale University

Susannah Fox; Pew Internet Project

Judy Freeman; Schroeder Institute for Tobacco Research & Policy Studies, Legacy

Gabe Garcia, PhD; University of Alaska Anchorage

Joe Gitchell; Pinney Associates, Inc.

Amanda Graham*, PhD; Schroeder Institute for Tobacco Research & Policy Studies, Legacy

Anne M. Hartman, MS, MA; National Cancer Institute, NIH

XSL•FO
RenderX

Colleen Haydon, MSW, MPH; CYAN/Project UNIFORM

Dave Heilmann; SparkPeople, Inc.

Thaddeus Herzog, PhD; Cancer Research Center of Hawaii

Kimberlee Homer Vagadori, MPH; California Youth Advocacy Network

Thomas Houston, MD; University of Massachusetts Medical School

Yvonne Hunt, PhD, MPH; National Cancer Institute, NIH

Kevin Hwang, MD, MPH; The University of Texas Medical School at Houston

Caroline Joyce; Legacy

Ross Kauffman, PhD; Indiana University

Katie Kemper, MBA; Mayo Clinic

Tom Kirchner, PhD; Schroeder Institute for Tobacco Research & Policy Studies, Legacy

Amy Knowlton, ScD; Johns Hopkins Bloomberg School of Public Health

Emily Z. Kontos, ScD; Harvard University, School of Public Health

Sanjay Koyani, MPH; FDA/Center for Tobacco Products

Janna Lacatell, MBA; Healthways Inc.

Stephanie R. Land, PhD; University of Pittsburgh and ReSET Center

Yuelin Li, PhD; Memorial Sloan-Kettering Cancer Center

Paula Lozano, MD, MPH; University of Washington

Patty Mabry*, PhD; Office of Behavioral and Social Sciences Research, NIH

Michael Macy, PhD; Cornell University

Stephen Marcus, PhD; National Institute of General Medical Sciences, NIH

Darren Mays, PhD; Georgetown University

Anna McDaniel, PhD, RN; Indiana University

Karen McDonnell, PhD, MSN, RN; University of Virginia

Howard Meitiner; Phoenix House

Robin Mermelstein*, PhD; University of Illinois at Chicago

Aaron Mushro; Legacy

Toben Nelson, ScD; University of Minnesota

Ray Niaura*, PhD; Schroeder Institute for Tobacco Research & Policy Studies, Legacy

Janet Okamoto, PhD; National Cancer Institute, NIH

George Papandonatos*, PhD; Brown University

Heather Patrick, PhD; National Cancer Institute, NIH

Pallavi Patwardhan, PhD; Schroeder Institute for Tobacco Research & Policy Studies, Legacy

Jennifer Pearson, MPH; Schroeder Institute for Tobacco Research & Policy Studies, Legacy

Alan Peters, CTTS-M; QuitNet / Healthways

Alison Pilsner, MPH, CPH, CHES; MMG, Inc.

Craig Pollack, MD; Johns Hopkins University

Danielle Ramo, PhD; University of California, San Francisco

Brandi Robinson, MPH; Partnership for Prevention

XSL•FO
RenderX

Heather Rogers, PhD, MPH; Concurrent Technologies Corporation

Marcel Salathé, PhD; Penn State University

Saul Shiffman*, PhD; University of Pittsburgh

Tom Snijders, PhD; University of Oxford

Karen Sodomick, MA; Phoenix House

Mike Spittel, PhD; National Institute of Child Health & Human Development, NIH

Bonnie Spring, PhD; Northwestern University

Cassandra Stanton*, PhD; The Warren Alpert Medical School of Brown University

Maggy Sterner; Small World consulting

Erin L. Sutfin, PhD; Wake Forest University School of Medicine

Samarth Swarup, PhD; Virginia Tech

Shani C. Taylor; MMG, Inc.

Hilary Tindle, MD, MPH; University of Pittsburgh

Tom Valente, PhD; University of Southern California Keck School of Medicine

Donna Vallone*, PhD, MPH; Legacy

Trevor van Mierlo, MScCH; Evolution Health Systems Inc.

Andrea Villanti, MPH, PhD(c); Schroeder Institute for Tobacco Research & Policy Studies, Legacy

J. Lee Westmaas, PHD; American Cancer Society

Robyn Whittaker, MD; University of Auckland

** Participant in the conference planning committee and/or post-workshop summary workgroup.*

Conflicts of Interest

Dr Cobb is a consultant to Healthways Inc., which operates QuitNet, a web-based smoking cessation application using social networks.

Multimedia Appendix 1

Conference introduction, Dr. David Abrams.

[MOV File, 24MB - jmir_v13i4e119_app1.mov]

Multimedia Appendix 2

Conference introduction - Dr. Saul Shiffman.

[MOV File, 32MB - jmir_v13i4e119_app2.mov]

Multimedia Appendix 3

Theme 1: Social support, health behavior and smoking cessation - Dr. Robin Mermelstein.

[MOV File, 65MB - jmir_v13i4e119_app3.mov]

Multimedia Appendix 4

Theme 1: Social support, health behavior and smoking cessation - Dr. Thomas Valente.

[MOV File, 56MB - jmir_v13i4e119_app4.mov]

Multimedia Appendix 5

Conference keynote - Dr. Nicholas Christakis.

XSL•FO
RenderX

[M4V File, 121MB - jmir_v13i4e119_app5.m4v]

Multimedia Appendix 6

Theme 2: Online social networks - Dr. Nathan Cobb.

[MOV File, 165MB - jmir_v13i4e119_app6.mov]

Multimedia Appendix 7

Theme 2: Online social networks - Dr. Noshir Contractor.

[MOV File, 153MB - jmir_v13i4e119_app7.mov]

Multimedia Appendix 8

Theme 2: Online social networks - Dr. Nathan Eagle.

[MOV File, 205MB - jmir_v13i4e119_app8.mov]

Multimedia Appendix 9

Theme 3: Intervention approaches - Mr. Dave Heilmann.

[MOV File, 114MB - jmir_v13i4e119_app9.mov]

Multimedia Appendix 10

Theme 3: Intervention approaches - Mr. Trevor va Mierlo.

[MOV File, 150MB - jmir_v13i4e119_app10.mov]

Multimedia Appendix 11

Theme 3: Intervention approaches - Mr. Chris Cartter.

[MOV File, 205MB - jmir_v13i4e119_app11.mov]

Multimedia Appendix 12

Theme 4: Methods, design and analysis - Dr. Linda Collins.

[M4V File, 173MB - jmir_v13i4e119_app12.m4v]

Multimedia Appendix 13

Theme 4: Methods, design and analysis - Dr. Tom Snijders.

[M4V File, 199MB - jmir_v13i4e119_app13.m4v]

References

1. Centers for Disease Control and Prevention (CDC). Smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 2000-2004. MMWR Morb Mortal Wkly Rep 2008 Nov 14;57(45):1226-1228 [FREE Full text] [Medline: 19008791]
2. Centers for Disease Control and Prevention (CDC). Cigarette smoking among adults and trends in smoking cessation - United States, 2008. MMWR Morb Mortal Wkly Rep 2009 Nov 13;58(44):1227-1232 [FREE Full text] [Medline: 19910909]
3. US Department of Health and Human Services (USDHHS). Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: US Department of Health and Human Services, Public Health Services; 2008.
4. Shiffman S, Brockwell SE, Pillitteri JL, Gitchell JG. Use of smoking-cessation treatments in the United States. Am J Prev Med 2008 Feb;34(2):102-111. [doi: 10.1016/j.amepre.2007.09.033] [Medline: 18201639]
5. Healthy People 2020. US Department of Health and Human Services. 2010 Dec 2. Tobacco Use Objectives URL: http://www.healthypeople.gov/2020/topicsobjectives2020/objectiveslist.aspx?topicId=41 [accessed 2011-06-22] [WebCite Cache ID 5zdgFH0ss]
6. Westmaas JL, Bontemps-Jones J, Bauer JE. Social support in smoking cessation: reconciling theory and evidence. Nicotine Tob Res 2010 Jul;12(7):695-707. [doi: 10.1093/ntr/ntq077] [Medline: 20513695]

XSL•FO
RenderX

7. Cohen S, Lemay EP. Why would social networks be linked to affect and health practices? Health Psychol 2007 Jul;26(4):410-417. [doi: 10.1037/0278-6133.26.4.410] [Medline: 17605560]
8. Christakis NA. Health care in a web. BMJ 2008 Jun 28;336(7659):1468. [doi: 10.1136/bmj.a452] [Medline: 18583678]
9. May S, West R. Do social support interventions ("buddy systems") aid smoking cessation? A review tobacco control. Tob Control 2000 Dec;9(4):415-422 [FREE Full text] [Medline: 11106712]
10. Lichtenstein E, Glasgow RE, Abrams DB. Social support in smoking cessation: in search of effective interventions. Behav Ther 1986;17:607-619. [doi: 10.1016/S0005-7894(86)80098-3]
11. Murray RP, Johnston JJ, Dolce JJ, Lee WW, O'Hara P. Social support for smoking cessation and abstinence: the Lung Health Study. Lung Health Study Research Group. Addict Behav 1995;20(2):159-170. [Medline: 7484310]
12. Park EW, Tudiver F, Schultz JK, Campbell T. Does enhancing partner support and interaction improve smoking cessation? A meta-analysis. Ann Fam Med 2004;2(2):170-174 [FREE Full text] [Medline: 15083859]
13. Graham AL, Cobb NK, Papandonatos GD, Moreno JL, Kang H, Tinkelman DG, et al. A randomized trial of Internet and telephone treatment for smoking cessation. Arch Intern Med 2011 Jan 10;171(1):46-53. [doi: 10.1001/archinternmed.2010.451] [Medline: 21220660]
14. Selby P, van Mierlo T, Voci SC, Parent D, Cunningham JA. Online social and professional support for smokers trying to quit: an exploration of first time posts from 2562 members. J Med Internet Res 2010;12(3):e34 [FREE Full text] [doi: 10.2196/jmir.1340] [Medline: 20719739]
15. Burri M, Baujard V, Etter JF. A qualitative analysis of an internet discussion forum for recent ex-smokers. Nicotine Tob Res 2006 Dec;8 Suppl 1:S13-S19. [Medline: 17491166]
16. Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. Am J Public Health 2010 Jul;100(7):1282-1289. [doi: 10.2105/AJPH.2009.165449] [Medline: 20466971]
17. comScore, Inc. 2008 Nov 19. Eighty Five Percent of Brazilian Internet Users Visited a Social Networking Site in September 2008 URL: http://www.comscore.com/Press_Events/Press_Releases/2008/11/Social_Networking_in_Brazil [accessed 2011-06-22] [WebCite Cache ID 5zdhC6oRb]
18. Facebook. 2010 Oct 1. Facebook press room URL: http://www.facebook.com/press/info.php?statistics [WebCite Cache ID 5tyQEzO17]
19. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. Addiction 2009 Nov;104(11):1792-1804. [doi: 10.1111/j.1360-0443.2009.02710.x] [Medline: 19832783]
20. Norman CD, McIntosh S, Selby P, Eysenbach G. Web-assisted tobacco interventions: empowering change in the global fight for the public's (e)Health. J Med Internet Res 2008;10(5):e48 [FREE Full text] [doi: 10.2196/jmir.1171] [Medline: 19033147]
21. Lazer D, Pentland A, Adamic L, Aral S, Barabasi AL, Brewer D, et al. Social science. Computational social science. Science 2009 Feb 6;323(5915):721-723. [doi: 10.1126/science.1167742] [Medline: 19197046]
22. Anonymous. Community cleverness required. Nature 2008;455(7209):1. [Medline: 18769385]
23. Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The spread of alcohol consumption behavior in a large social network. Ann Intern Med 2010 Apr 6;152(7):426-33, W141. [doi: 10.1059/0003-4819-152-7-201004060-00007] [Medline: 20368648]
24. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379 [FREE Full text] [doi: 10.1056/NEJMsa066082] [Medline: 17652652]
25. Hammond RA. Social influence and obesity. Curr Opin Endocrinol Diabetes Obes 2010 Oct;17(5):467-471. [doi: 10.1097/MED.0b013e32833d4687] [Medline: 20689421]
26. Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. BMJ 2008;337:a2338 [FREE Full text] [Medline: 19056788]
27. Rosenquist JN, Fowler JH, Christakis NA. Social network determinants of depression. Mol Psychiatry 2011 Mar;16(3):273-281. [doi: 10.1038/mp.2010.13] [Medline: 20231839]
28. Cacioppo JT, Fowler JH, Christakis NA. Alone in the crowd: the structure and spread of loneliness in a large social network. J Pers Soc Psychol 2009 Dec;97(6):977-991. [doi: 10.1037/a0016076] [Medline: 19968414]
29. Moreno JL. Who Shall Survive? A New Approach to the Problem of Human Interrelations. Washington, DC: Nervous and Mental Disease Publishing Co; 1934.
30. Mabry PL, Olster DH, Morgan GD, Abrams DB. Interdisciplinarity and systems science to improve population health: a view from the NIH Office of Behavioral and Social Sciences Research. Am J Prev Med 2008 Aug;35(2 Suppl):S211-S224. [doi: 10.1016/j.amepre.2008.05.018] [Medline: 18619402]
31. Graham AL, Abrams DB. Reducing the cancer burden of lifestyle factors: opportunities and challenges of the Internet. J Med Internet Res 2005 Jul 1;7(3):e26 [FREE Full text] [doi: 10.2196/jmir.7.3.e26] [Medline: 15998617]
32. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.
33. Mermelstein R, Cohen S, Lichtenstein E, Baer JS, Kamarck T. Social support and smoking cessation and maintenance. J Consult Clin Psychol 1986 Aug;54(4):447-453. [Medline: 3745596]
34. Steglich C, Snijders TAB, Pearson M. Dynamic networks and behavior: separating selection from influence. Sociol Methodol 2010 Aug;40(1):329-393. [doi: 10.1111/j.1467-9531.2010.01225.x]

XSL•FO
RenderX

35. Simmel G. On Individuality and Social Forms: Selected Writings. Chicago, IL: University of Chicago Press; 1971.
36. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. Am J Epidemiol 1979 Feb;109(2):186-204. [Medline: 425958]
37. Berkman LF, Glass T. Social integration, social networks, social support, and health. In: Berkman LF, Kawachi I, editors. Social Epidemiology. New York, NY: Oxford University Press; 2000:137-173.
38. Paul MJ, Dredze M. You are what you tweet: analyzing twitter for public health. In: Proceedings. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media (ICWSM); July 17-21, 2011; Barcelona, Spain.
39. Centola D. The spread of behavior in an online social network experiment. Science 2010 Sep 3;329(5996):1194-1197 [FREE Full text] [doi: 10.1126/science.1185231] [Medline: 20813952]
40. Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proc Natl Acad Sci U S A 2009 Dec 22;106(51):21544-21549 [FREE Full text] [doi: 10.1073/pnas.0908800106] [Medline: 20007780]
41. Salganik MJ, Dodds PS, Watts DJ. Experimental study of inequality and unpredictability in an artificial cultural market. Science 2006 Feb 10;311(5762):854-856 [FREE Full text] [doi: 10.1126/science.1121066] [Medline: 16469928]
42. Kossinets G, Watts DJ. Empirical analysis of an evolving social network. Science 2006 Jan 6;311(5757):88-90 [FREE Full text] [doi: 10.1126/science.1116869] [Medline: 16400149]
43. Eagle N, Macy M, Claxton R. Network diversity and economic development. Science 2010 May 21;328(5981):1029-1031 [FREE Full text] [doi: 10.1126/science.1186605] [Medline: 20489022]
44. Newman MEJ. The structure and function of complex networks. SIAM Rev 2003;45(2):167-256. [doi: 10.1137/S003614450342480]
45. Abrams DB, Leslie F, Mermelstein R, Kobus K, Clayton RR. Transdisciplinary tobacco use research. Nicotine Tob Res 2003 Dec;5 Suppl 1:S5-10. [doi: 10.1080/14622200310001625519] [Medline: 14668083]
46. Marin A, Wellman B. Social network analysis: an introduction. In: Scott PJP, Carrington PP, editors. Scott J, Carrington PJ, editors. The SAGE Handbook of Social Network Analysis. Thousand Oaks, CA: SAGE Publications; 2011:11-25.
47. Bock BC, Graham AL, Whiteley JA, Stoddard JL. A review of web-assisted tobacco interventions (WATIs). J Med Internet Res 2008;10(5):e39 [FREE Full text] [doi: 10.2196/jmir.989] [Medline: 19000979]
48. Weitzman ER, Cole E, Kaci L, Mandl KD. Social but safe? Quality and safety of diabetes-related online social networks. J Am Med Inform Assoc 2011 May 1;18(3):292-297. [doi: 10.1136/jamia.2010.009712] [Medline: 21262920]
49. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]
50. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Ann Rev Sociol 2001;27:415-444. [doi: 10.1146/annurev.soc.27.1.415]
51. Rogers EM. Diffusion of Innovations. New York, NY: Free Press; 2003.
52. Iyengar R, Van den Bulte C, Valente TW. Opinion leadership and social contagion in new product diffusion. Marketing Sci 2010;30(2):195-212. [doi: 10.1287/mksc.1100.0566]
53. Watts DJ, Peretti J. Viral marketing for the real world. Harvard Bus Rev 2007;85(5):22-23.
54. Lessig L. Code and Other Laws of Cyberspace. New York, NY: Basic Books; 1999.
55. Lampe CAC, Ellison N, Steinfield C. A familiar face(book): profile elements as signals in an online social network. In: Proceedings. 2007 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 28-May 3, 2007; San Jose, CA, USA.
56. Bansal MA, Cummings KM, Hyland A, Giovino GA. Stop-smoking medications: who uses them, who misuses them, and who is misinformed about them? Nicotine Tob Res 2004 Dec;6 Suppl 3:S303-S310. [Medline: 15799593]
57. An LC, Schillo BA, Saul JE, Wendling AH, Klatt CM, Berg CJ, et al. Utilization of smoking cessation informational, interactive, and online community resources as predictors of abstinence: cohort study. J Med Internet Res 2008;10(5):e55 [FREE Full text] [doi: 10.2196/jmir.1018] [Medline: 19103587]
58. Stoddard JL, Augustson EM, Moser RP. Effect of adding a virtual community (bulletin board) to smokefree.gov: randomized controlled trial. J Med Internet Res 2008;10(5):e53 [FREE Full text] [doi: 10.2196/jmir.1124] [Medline: 19097974]
59. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
60. Cobb NK, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. Nicotine Tob Res 2005 Apr;7(2):207-216. [doi: 10.1080/14622200500055319] [Medline: 16036277]
61. Graham AL, Bock BC, Cobb NK, Niaura R, Abrams DB. Characteristics of smokers reached and recruited to an internet smoking cessation trial: a case of denominators. Nicotine Tob Res 2006 Dec;8 Suppl 1:S43-S48. [Medline: 17491170]
62. 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] [doi: 10.2196/jmir.7.1.e2] [Medline: 15829474]
63. Cobb NK, Graham AL. Characterizing Internet searchers of smoking cessation information. J Med Internet Res 2006;8(3):e17 [FREE Full text] [doi: 10.2196/jmir.8.3.e17] [Medline: 17032633]

XSL•FO
RenderX

64. Vallone DM, Duke JC, Cullen J, McCausland KL, Allen JA. Evaluation of EX: a national mass media smoking cessation campaign. Am J Public Health 2011 Feb;101(2):302-309. [doi: 10.2105/AJPH.2009.190454] [Medline: 21164094]
65. Bala M, Strzeszynski L, Cahill K. Mass media interventions for smoking cessation in adults. Cochrane Database Syst Rev 2008(1):CD004704. [doi: 10.1002/14651858.CD004704.pub2] [Medline: 18254058]
66. Abrams DB, Mills S, Bulger D. Challenges and future directions for tailored communication research. Ann Behav Med 1999;21(4):299-306. [Medline: 10721436]
67. Strecher VJ, McClure JB, Alexander GL, Chakraborty B, Nair VN, Konkel JM, et al. Web-based smoking-cessation programs: results of a randomized trial. Am J Prev Med 2008 May;34(5):373-381. [doi: 10.1016/j.amepre.2007.12.024] [Medline: 18407003]
68. Zbikowski SM, Hapgood J, Smucker Barnwell S, McAfee T. Phone and web-based tobacco cessation treatment: real-world utilization patterns and outcomes for 11,000 tobacco users. J Med Internet Res 2008;10(5):e41 [FREE Full text] [doi: 10.2196/jmir.999] [Medline: 19017583]
69. Zbikowski SM, Jack LM, McClure JB, Deprey M, Javitz HS, McAfee TA, et al. Utilization of services in a randomized trial testing phone- and web-based interventions for smoking cessation. Nicotine Tob Res 2011 May;13(5):319-327. [doi: 10.1093/ntr/ntq257] [Medline: 21330267]
70. Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, Smith SS, et al. The multiphase optimization strategy for engineering effective tobacco use interventions. Ann Behav Med 2011 Apr;41(2):208-226. [doi: 10.1007/s12160-010-9253-x] [Medline: 21132416]
71. Pike KJ, Rabius V, McAlister A, Geiger A. American Cancer Society's QuitLink: randomized trial of Internet assistance. Nicotine Tob Res 2007 Mar;9(3):415-420. [doi: 10.1080/14622200701188877] [Medline: 17365773]
72. Strecher VJ, Shiffman S, West R. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. Addiction 2005 May;100(5):682-688. [doi: 10.1111/j.1360-0443.2005.01093.x] [Medline: 15847626]
73. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010(9):CD007078. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
74. Swartz LH, Noell JW, Schroeder SW, Ary DV. A randomised control study of a fully automated internet based smoking cessation programme. Tob Control 2006 Feb;15(1):7-12 [FREE Full text] [doi: 10.1136/tc.2003.006189] [Medline: 16436397]
75. Muñoz RF, Barrera AZ, Delucchi K, Penilla C, Torres LD, Pérez-Stable EJ. International Spanish/English Internet smoking cessation trial yields 20% abstinence rates at 1 year. Nicotine Tob Res 2009 Sep;11(9):1025-1034. [doi: 10.1093/ntr/ntp090] [Medline: 19640833]
76. Japuntich SJ, Zehner ME, Smith SS, Jorenby DE, Valdez JA, Fiore MC, et al. Smoking cessation via the internet: a randomized clinical trial of an internet intervention as adjuvant treatment in a smoking cessation intervention. Nicotine Tob Res 2006 Dec;8 Suppl 1:S59-S67. [Medline: 17491172]
77. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39. [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]
78. Eysenbach G. Issues in evaluating health websites in an Internet-based randomized controlled trial. J Med Internet Res 2002 Dec;4(3):E17 [FREE Full text] [doi: 10.2196/jmir.4.3.e17] [Medline: 12554548]
79. Glasgow RE. Enhancing the scientific foundation of internet intervention research. Ann Behav Med 2009 Aug;38(1):46-47. [doi: 10.1007/s12160-009-9134-3] [Medline: 19816749]
80. Ahern DK. Challenges and opportunities of eHealth research. Am J Prev Med 2007 May;32(5 Suppl):S75-S82. [doi: 10.1016/j.amepre.2007.01.016] [Medline: 17466822]
81. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
82. Eagle N, Pentland AS, Lazer D. Inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci U S A 2009 Sep 8;106(36):15274-15278 [FREE Full text] [doi: 10.1073/pnas.0900282106] [Medline: 19706491]
83. Christakis NA. Social networks and collateral health effects. BMJ 2004 Jul 24;329(7459):184-185 [FREE Full text] [doi: 10.1136/bmj.329.7459.184] [Medline: 15271805]
84. Galea S, Riddle M, Kaplan GA. Causal thinking and complex system approaches in epidemiology. Int J Epidemiol 2010 Feb;39(1):97-106 [FREE Full text] [doi: 10.1093/ije/dyp296] [Medline: 19820105]
85. Levy DT, Mabry PL, Graham AL, Orleans CT, Abrams DB. Reaching Healthy People 2010 by 2013: A SimSmoke simulation. Am J Prev Med 2010 Mar;38(3 Suppl):S373-S381. [doi: 10.1016/j.amepre.2009.11.018] [Medline: 20176310]
86. Killeen PR. Markov model of smoking cessation. Proc Natl Acad Sci U S A 2011 Sep 13;108 Suppl 3:15549-15556. [doi: 10.1073/pnas.1011277108] [Medline: 21508324]
87. Abrams DB. Comprehensive smoking cessation policy for all smokers: systems integration to save lives and money. In: Bonnie RJ, Stratton KR, Wallace RB, editors. Ending the Tobacco Problem: A Blueprint for the Nation . Washington, DC: National Academies Press; 2007.
88. Abrams DB, Graham AL, Levy DT, Mabry PL, Orleans CT. Boosting population quits through evidence-based cessation treatment and policy. Am J Prev Med 2010 Mar;38(3 Suppl):S351-S363. [doi: 10.1016/j.amepre.2009.12.011] [Medline: 20176308]

XSL•FO
RenderX

89. Levy DT, Graham AL, Mabry PL, Abrams DB, Orleans CT. Modeling the impact of smoking-cessation treatment policies on quit rates. Am J Prev Med 2010 Mar;38(3 Suppl):S364-S372. [doi: 10.1016/j.amepre.2009.11.016] [Medline: 20176309]
90. Morell JA, Hilscher R, Magura S, Ford J. Integrating evaluation and agent-based modeling: rationale and an example for adopting evidence-based practices. J Multidiscip Eval 2010 Aug;6(14):32-57.

Edited by G Eysenbach; submitted 13.08.11; peer-reviewed by P Wicks; comments to author 27.08.11; revised version received 19.09.11; accepted 25.09.11; published 19.12.11.

Please cite as:
Cobb NK, Graham AL, Byron MJ, Niaura RS, Abrams DB, Workshop Participants
Online Social Networks and Smoking Cessation: A Scientific Research Agenda
J Med Internet Res 2011;13(4):e119
URL: http://www.jmir.org/2011/4/e119/
doi:10.2196/jmir.1911
PMID:22182518

©Nathan K Cobb, Amanda L Graham, M Justin Byron, David B Abrams, Workshop Participants. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 19.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Anxiety Online—A Virtual Clinic: Preliminary Outcomes Following Completion of Five Fully Automated Treatment Programs for Anxiety Disorders and Symptoms

Britt Klein[1,2,3], DPsych (Clinical); Denny Meyer[1,2,3], DBL; David William Austin[1,2,3], PhD; Michael Kyrios[1,2,3], PhD

[1]National eTherapy Centre, Faculty of Life and Social Sciences, Swinburne University, Melbourne, Australia

[2]Brain and Psychological Sciences Research Centre, Faculty of Life and Social Sciences, Swinburne University, Melbourne, Australia

[3]eTherapy Research Unit, Faculty of Life and Social Sciences, Swinburne University, Melbourne, Australia

Corresponding Author:
Britt Klein, DPsych (Clinical)
National eTherapy Centre
Faculty of Life and Social Sciences
Swinburne University
Mail H99
PO Box 218, Hawthorn
Melbourne, 3122
Australia
Phone: 61 392148851
Fax: 61 392145260
Email: bklein@swin.edu.au

Abstract

Background: The development of e-mental health interventions to treat or prevent mental illness and to enhance wellbeing has risen rapidly over the past decade. This development assists the public in sidestepping some of the obstacles that are often encountered when trying to access traditional face-to-face mental health care services.

Objective: The objective of our study was to investigate the posttreatment effectiveness of five fully automated self-help cognitive behavior e-therapy programs for generalized anxiety disorder (GAD), panic disorder with or without agoraphobia (PD/A), obsessive–compulsive disorder (OCD), posttraumatic stress disorder (PTSD), and social anxiety disorder (SAD) offered to the international public via Anxiety Online, an open-access full-service virtual psychology clinic for anxiety disorders.

Methods: We used a naturalistic participant choice, quasi-experimental design to evaluate each of the five Anxiety Online fully automated self-help e-therapy programs. Participants were required to have at least subclinical levels of one of the anxiety disorders to be offered the associated disorder-specific fully automated self-help e-therapy program. These programs are offered free of charge via Anxiety Online.

Results: A total of 225 people self-selected one of the five e-therapy programs (GAD, n = 88; SAD, n = 50; PD/A, n = 40; PTSD, n = 30; OCD, n = 17) and completed their 12-week posttreatment assessment. Significant improvements were found on 21/25 measures across the five fully automated self-help programs. At postassessment we observed significant reductions on all five anxiety disorder clinical disorder severity ratings (Cohen d range 0.72–1.22), increased confidence in managing one's own mental health care (Cohen d range 0.70–1.17), and decreases in the total number of clinical diagnoses (except for the PD/A program, where a positive trend was found) (Cohen d range 0.45–1.08). In addition, we found significant improvements in quality of life for the GAD, OCD, PTSD, and SAD e-therapy programs (Cohen d range 0.11–0.96) and significant reductions relating to general psychological distress levels for the GAD, PD/A, and PTSD e-therapy programs (Cohen d range 0.23–1.16). Overall, treatment satisfaction was good across all five e-therapy programs, and posttreatment assessment completers reported using their e-therapy program an average of 395.60 (SD 272.2) minutes over the 12-week treatment period.

Conclusions: Overall, all five fully automated self-help e-therapy programs appear to be delivering promising high-quality outcomes; however, the results require replication.

Trial Registration: Australian and New Zealand Clinical Trials Registry ACTRN121611000704998; http://www.anzctr.org.au/trial_view.aspx?ID=336143 (Archived by WebCite at http://www.webcitation.org/618r3wvOG)

XSL•FO
RenderX

(J Med Internet Res 2011;13(4):e89) doi:10.2196/jmir.1918

KEYWORDS

eTherapy; Internet interventions; e-mental health; cognitive behavior therapy; generalized anxiety disorder; panic disorder; obsessive compulsive disorder; social anxiety disorder; posttraumatic stress disorder; self-help; fully automated; Web treatment

Introduction

Anxiety disorder is a generic term given to a group of specific disorders that are typically characterized by fear, worry, and phobic responses. The main anxiety disorder types are generalized anxiety disorder (GAD), panic disorder with or without agoraphobia (PD/A), obsessive–compulsive disorder (OCD), posttraumatic stress disorder (PTSD), and social anxiety disorder (SAD). These disorders are highly prevalent mental health conditions that have deleterious effects on a person's life, including substantial personal, social, and occupational impairment, and are often associated with considerable comorbidity [1-4] resulting in significant economic costs for the individual and society. In the Australian National Mental Health Survey [2], only around one-third of those with an anxiety disorder (37.8%) reported making use of services over the previous 12 months for their mental health problems [5].

Cognitive behavior therapy (CBT) has been shown to be an effective treatment for GAD, PD/A, OCD, PSTD, and SAD. Face-to-face CBT for these anxiety disorders typically involves 60 to 90 minutes of treatment per week over 9–12 weeks, including psychoeducation, anxiety management (eg, relaxation techniques), cognitive and exposure therapy, and relapse prevention [6]. Nevertheless, this form of specialized treatment is unavailable to many of those affected due to a shortage of suitably qualified health care professionals (especially in regional and rural areas), fee-for-service costs, and the stigma attached to seeing a mental health professional [7,8].

The development of e-therapy or e-mental health interventions (delivery of mental health interventions and services via information and communication technologies) has grown exponentially over the past decade, and is one way of delivering CBT that overcomes the commonly cited obstacles to treatment provision [9]. There are now hundreds of e-mental health interventions designed to treat or prevent mental illness and to enhance well-being. A helpful practitioner and consumer resource that provides information and quality ratings for over 180 e-physical health and e-mental health interventions can be accessed via Beacon [10], an online portal to eHealth interventions [11].

The most common type of e-mental health intervention is the Internet- or Web-based intervention or e-therapy (see [12]). e-Therapy programs can be broadly categorized as being self-help or therapist-assisted, and hundreds have been evaluated across a range of mental health disorders and symptoms, including panic disorder (eg, [13,14], SAD (eg, [15]), PTSD and symptoms (eg, [8,16,17]), anxiety prevention (eg, [18,19]), depression and depressive symptoms (eg, [20-24]), insomnia (eg, [25]), and alcohol issues (eg, [26]). Additionally, therapist-assisted e-therapy treatment programs have been found to be as effective as best-practice face-to-face therapy [13,27]. Numerous reviews [28-30] and meta-analyses (eg, [31,32]) attest to the general effectiveness of e-therapies based on validated therapeutic models such as CBT.

Although hundreds of e-therapy programs have been developed, the vast majority are generally accessible only via participation in research trials. In addition, most of the programs developed are singular offerings rather than broad-based virtual clinics offering multiple services. However, several groups offer an array of e-therapy programs contained within a single platform, such as e-hub [33], eCentreClinic [34], and Anxiety Online [35]. e-hub, operating through the Australian National University, Canberra, Australia, provides a variety of open-access self-help programs for mental health and well-being, such as MoodGYM, BluePages, BlueBoard, and e-couch [36], to the worldwide public. However, these e-mental health programs were largely designed to prevent ill health, rather than to treat clinical populations, and online therapist assistance is not offered (although BlueBoard, an Internet support group facility, includes human moderators who oversee consumer postings and appropriate online behaviors). On the other hand, the eCentreClinic, operating through Macquarie University, Sydney, Australia, offers a range of self-help and therapist-assisted e-therapy treatment programs for the anxiety disorders and depression. However, access is restricted to participation in research trials, opened only to the Australian public and at different times during the year.

Anxiety Online, operating through the National eTherapy Centre at Swinburne University of Technology, and funded by the Australian Government Department of Health and Ageing, provides to the international public a full-service, open-access, virtual psychology clinic for anxiety disorders. More specifically, Anxiety Online comprises four major components: (1) an open-access psychoeducational website that provides information about Anxiety Online, anxiety disorders (symptoms, prevalence, how and where they are treated), links to useful resources, and an entry/registration point for consumers, health care practitioners, and administrators, (2) a freely available online psychological assessment and referral system (e-PASS) that assesses the person for symptoms associated with 21 disorders in the Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR, 4th edition, text revision (DSM-IV-TR) [37] (ie, PD/A, agoraphobia without history of panic disorder, SAD, specific phobia, GAD, PTSD, OCD, depression, anorexia nervosa, bulimia nervosa, binge eating disorder, somatization, body dysmorphic disorder, pathological gambling, insomnia, hypersomnia, alcohol dependence, and substance dependence—cannabis, opioids, sedatives, and stimulants), (3) five interactive, fully automated, 12-module self-help or therapist-assisted (via email) e-therapy treatment programs for GAD, PD/A, PTSD, OCD, and SAD, and (4) online e-therapist/CBT training programs and a health care practitioner portal.

XSL•FO
RenderX

Before commencing work as an e-therapist at the National eTherapy Centre, all prospective therapists are provided online e-therapy and CBT training and must pass a competency-based assessment (see [38] for more details). Many of the Anxiety Online e-therapists are postgraduate psychology students from various Australian universities who are undertaking an e-therapy psychological internship or placement. At a minimum, all have provisional registration as a psychologist. Anxiety Online also provides health care professionals worldwide (eg, general practitioners, psychologists, social workers, mental health nurses, aboriginal health workers, and psychiatrists) free access to the Anxiety Online programs (visit [39] to register).

Anxiety Online was launched for the international public in October 2009. This paper reports on the pre- to posttreatment outcome results for the completers of the five fully automated self-help treatment programs from October 2009 to April 2011. The goal of the Anxiety Online service is to increase access to mental health services by reducing the common obstacles and to provide consumers with choice in regard to treatment, as is the case for real-world settings. Therefore, we used a naturalistic design to evaluate the mental health treatment outcomes. The primary treatment outcome measure was the anxiety disorder severity ratings, with secondary outcome measures relating to general psychological distress levels, total number of DSM-IV-TR [37] diagnoses, confidence in managing one's own mental health care, and quality of life. We expected that after completing one of the five Anxiety Online treatment programs, participants would show decreases in their anxiety disorder severity rating, general psychological distress levels, and the total number of mental health diagnoses at posttreatment, as well as improvements in confidence in managing their own mental health care and quality of life.

Methods

Participants and Flow

Anxiety Online is an open-access website platform. We recruit participants via periodic Facebook advertisements, referral links on other mental health websites, use of local and national media, and presentations and brochure mail-outs to health care practitioners and consumer groups.

When visiting Anxiety Online, participants wanting to undertake e-PASS are first required to register and consent to the Anxiety Online terms and conditions [40]. After providing consent, participants are taken to e-PASS, which is the gateway to the fully automated self-help and therapist-assisted treatment programs. e-PASS was designed to ensure that all participants were offered an appropriate e-therapy treatment program based on their reported symptoms, as well as a way to help them identify whether they are experiencing difficulties within a range of psychological symptoms and disorders. In addition to addressing 21 DSM-IV-TR [37] disorders, a variety of demographic and personal information (eg, whether they are currently accessing mental health treatment) is recorded.

The key inclusion criteria for access to the treatment programs are being 18 years of age or older, completing e-PASS, and having either a subclinical or clinical diagnosis of at least one of GAD, PD/A, OCD, PTSD, or SAD. From October 2009 until April 2011, there were 7140 legitimate e-PASS pretreatment completions. The Anxiety Online data file initially contained 7245 completed pretreatment e-PASS administrations; however, 105 were removed (ie, 81 reported being under 18 years of age; and 24 were identified as health care professionals or researchers not interested in using Anxiety Online for their own personal purposes). From the 7140 e-PASS pretreatment completers, 168 did not have any disorder or symptoms and an additional 593 did not have a clinical or subclinical diagnosis of GAD, PD/A, OCD, PTSD, or SAD, for which Anxiety Online has treatment programs. This left 6379 consumers being offered an Anxiety Online treatment program. From this, 2660 elected to start a program and 3719 elected not to. The overall Anxiety Online e-therapy program acceptance rate was therefore 42% (2660/6379). It is important to note, however, that only 2986 of 6379 participants had an anxiety disorder as their primary diagnosis and, therefore, 89% (2660/2986) of participants with a primary anxiety disorder elected to commence one of the e-therapy programs. This is important because e-PASS strongly encourages participants, via their e-PASS feedback report, to seek treatment for their primary condition first.

Of the 2660 who started an e-therapy program, 75 elected to take the therapist-assisted version (due to the small numbers, these data are not presented in this paper). Of the remaining 2585, at time of data analyses 350 of the participants in the fully automated self-help program were still in progress; thus, the total number of participants who had completed their 12-week treatment period was 2235. From this sample, 832 commenced GAD Online, 406 commenced Panic Stop!, 168 commenced OCD Stop!, 227 commenced PSTD Online, and 602 commenced SAD Online. Anxiety Online also collects e-PASS data from registered program users every year for 5 years, and this follow-up data will be reported in due course. The procedures for reporting of the Anxiety Online data were approved by the Swinburne University Human Research Ethics Committee. Trial registration was obtained retrospectively because Anxiety Online is an ongoing open-access mental health service rather than a pure research trial. The program automatically collects consumer data regarding treatment outcomes via e-PASS and therefore differs from the traditional trial study design that has a clear start and end date.

Measures

Assessment included an online automated self-report clinical interview (e-PASS) assessing for 21 DSM-IV-TR [37] disorders, plus several other online questionnaires at pre- or posttreatment, or both.

Online Psychological Assessment and Referral System

e-PASS (B Klein, DPsych (Clinical), unpublished data, September 2010) is an online self-report diagnostic tool that assesses for 21 DSM-IV-TR [37] disorders and serves as the gateway into the fully automated self-help and therapist-assisted treatment programs. Although e-PASS can be completed over several sittings, it must be completed within a 24-hour period. e-PASS is automated and consists of over 540 items directly using the criteria specified in the DSM-IV-TR [37]. In addition to addressing 21 DSM-IV-TR [37] disorders, a variety of

XSL•FO
RenderX

demographic (and personal) questions are asked, as well as several screening items (eg, suicide, distorted thinking). It also checks for whether medical conditions and substance affects may better account for reported symptoms. For those who report suicidal ideation or distorted thinking patterns, e-PASS strongly recommends that the test taker cease completing e-PASS and contact a more appropriate service (referral sources are provided).

e-PASS is a complicated system, using both a categorical and dimensional approach to diagnosis, as well as branching logic and algorithmic scoring rules to minimize the number of irrelevant items presented, and is sensitive to other possible causes for symptoms being reported (eg, medical conditions). As a result, the number and types of items presented differ depending on the symptoms being endorsed and this, in turn, affects the feedback provided to the participant via the comprehensive personalized report generated on completion of e-PASS.

The e-PASS feedback report includes likely primary diagnosis, any likely secondary diagnoses (ranked according to number of symptoms and self-reported severity), and whether each likely diagnosis is within a clinical or subclinical range. Disorders in the subclinical range refer to those individuals who report most, but not all, of the required DSM-IV-TR criteria or, alternatively, report all of the necessary DSM-IV-TR criteria but provide low distress and interference ratings regarding their specific disorder symptoms. People with subclinical disorders are symptomatic (or subthreshold) but do not meet full DSM-IV-TR criteria for a clinical disorder. Disorders at the clinical level are further defined as mild, moderate, or severe, and participants are given hyperlinked or pop-up information explaining in plain language what each of these terms means. Likely clinical disorder severity ratings range from 0 (absence of any symptoms) to 8 (very severe clinical disorder) and constitute one of the main outcome variables for this study. Likely clinical disorder severity scores below 3.50 are given a subclinical label and rating. Feedback reports strongly recommend that consumers address their primary condition first, but ultimately what course of action or treatment participants undertake remains their choice (ie, Anxiety Online enables access to treatment programs for each consumer's primary diagnosis and any secondary diagnoses). Participants are also provided with a summary of the symptoms for each condition they have, and qualifiers are provided where appropriate (eg, chronicity of PTSD). Consumers are also provided a recommended course of action and multiple referral options. As individuals remain completely free to choose whatever course of action they desire, within the confines of the symptoms reported, Anxiety Online is a participant choice-based system. Nevertheless, it also provides the participant with detailed and evidence-based guidance and recommendations.

e-PASS is undergoing psychometric validation and qualitative evaluation, and the pilot and preliminary data suggest it is an acceptable and valid diagnostic tool (B Klein, DPsych (Clinical), unpublished data, September 2010, [41]), although caution is still warranted until the full and detailed study is published. Basic community-based validation results also attest to its validity. Using the results from the current study, at posttreatment 64 participants reported that they sought confirmation of their e-PASS diagnoses with an external source (n = 33 with a psychologist, n = 16 with a medical doctor, n = 6 with a website, n = 4 with a counselor, n = 4 with a friend, and n = 1 with a book). The overall agreement rate was 95% (61/64). The three sources where agreement was not reached were a medical doctor (n = 1), a website (n = 1), and a friend (n = 1).

Online Questions/Questionnaire: Self-Report

Kessler-6 (K6) [42] is a brief 6-item self-report measure, using a Likert-type scale from 1 to 5, measuring nonspecific psychological distress over the last 30 days [42]. Scores range from 6 to 30, with higher scores indicating greater psychological distress. Normative data (G Andrews, MD, written communication, August 2010) suggest that scores between 6 and 11 indicate low distress levels (71.7% of the population); scores between 12 and 15 indicate moderate distress levels (16.6% of the population); scores between 16 and 19 indicate high distress levels (7.16% of the population); and scores between 20 and 30 indicate very high distress levels (2.5% of the population). The K6 has demonstrated strong psychometric properties (eg, [42,43])

Number of e-PASS diagnoses is the total number of e-PASS diagnoses (at clinical and subclinical levels) as assessed by e-PASS at pre- and posttreatment assessment.

Confidence in managing mental health is a single-question self-report item asking participants to rate their overall level of self-confidence in managing their own mental health. Scores are anchored (1 = very poor, 2 = poor, 3 = neither poor nor good, 4 = good, 5 = very good), with higher scores indicating greater self-confidence.

Quality of life is a single-item self-report question asking participants to rate their overall quality of life. Scores are anchored (1 = very poor, 2 = poor, 3 = neither poor nor good, 4 = good, 5 = very good), with higher scores indicating a higher quality of life.

The two *e-Therapy treatment satisfaction questions* ask participants to rate (1) how satisfied they were with the e-therapy program, using a scale from 1 (not at all) to 5 (very highly), and (2) how much they liked the e-therapy treatment program, using a scale of 0 (not at all), 2 (a little), 4 (somewhat), 6 (quite a lot), and 8 (very much so) at posttreatment.

Design

The five fully automated self-help e-therapy treatment programs were trialed using a pre- to posttreatment quasi-experimental (participant choice) naturalistic design. The five programs all have a similar structure and look. Each program addresses a particular anxiety disorder (ie, *GAD Online* treats GAD, *Panic Stop!* treats PD/A, *OCD Stop!* treats OCD, *PTSD Online* treats PTSD, and *SAD Online* treats SAD). Each program is based on well-established CBT principles and protocols and was reviewed by national and international experts. In addition, all programs were subjected to rigorous technical and consumer usability testing prior to launch.

XSL•FO
RenderX

All programs consist of 12 modules, delivered over 12 weeks, that include a variety of text-based and multimedia materials (audio, video, and animated graphics) and online activities—for example, video (expert speaking, patient speaking, examples of therapy techniques or sessions, etc), audio (breathing control, visual imagery, progressive relaxation therapy instructions, etc), online activities (weekly self-monitoring, quizzes, journal writing, etc), downloadable PDFs (worksheets, transcripts of the audio, monitoring forms, etc), and online interactive animations (flash animations to convey key concepts) (see Multimedia Appendix 1 for several screenshot examples). In addition, there were numerous automated emails welcoming participants to the program, reminding and encouraging them to log on and complete their assessments, as well as various "alert" emails that are triggered depending on participants' online behavior (eg, alert automated emails are triggered when participants' self-monitoring of their anxiety and depression remain static for 4 weeks, or remains in the upper extreme range for 2 weeks in a row). These automated emails alert the person to a particular issue and provide recommendations (eg, to consider seeking more intensive assistance). All e-therapy programs contain standard CBT content teachings with regard to psychoeducation, anxiety management, and physiological, cognitive, and behavioral change strategies specific to each anxiety disorder, as well as weekly online and offline homework activities.

Procedures

Participants self-register to use the Anxiety Online virtual clinic. All participants are required to read and agree to the terms and conditions of the Anxiety Online service before being able to proceed to e-PASS. If participants meet criteria (18 years of age and over, and receiving a diagnosis of GAD, PD/A, OCD, PTSD, or SAD), they are offered the e-therapy program(s) that treat their specific disorder(s).

For this study, once participants chose a fully automated self-help e-therapy treatment program, they gained immediate access to the program and their 12-week e-therapy treatment cycle commenced. During this 12-week period, participants could not undertake another e-therapy program that they may have been offered in their e-PASS report (participants are unable to do more than one e-therapy program concurrently). If participants no longer wanted to continue, they were required to opt out of their e-therapy program via an opt-out option provided within each e-therapy program.

After completing the e-therapy program (at the end of week 12), participants who had not opted out were sent automated emails asking them to complete their posttreatment assessment questions and e-PASS. Several reminder emails were sent out over a 3-week period to those who had not completed the posttreatment assessment in a timely manner.

Statistical Procedures, Analyses, and Evaluation of Treatment Effects

After multivariate analysis of variance (MANOVA) tests showed no significant attrition bias, treatment effects from pre- to postassessment were evaluated separately for each of the five fully automated self-help e-therapy treatment programs, using a repeated measures MANOVA for the five treatment outcome variables (ie, clinical disorder severity rating, K6 scores, number of diagnoses, confidence in managing one's own mental health care, and quality of life). Follow-up analysis of variance (ANOVA) tests were then conducted for the programs with significant results. Normality and homogeneity assumptions were supported by the data and effect sizes were established using Cohen's [44] classification scheme (small effect = 0.20, medium effect = 0.50, and large effect = 0.80). The 95% confidence intervals for the expected program changes are also presented. In addition, the e-therapy treatment satisfaction results are presented. We used SPSS version 19 for Windows (IBM Corporation, Somers, NY, USA) to analyze all data.

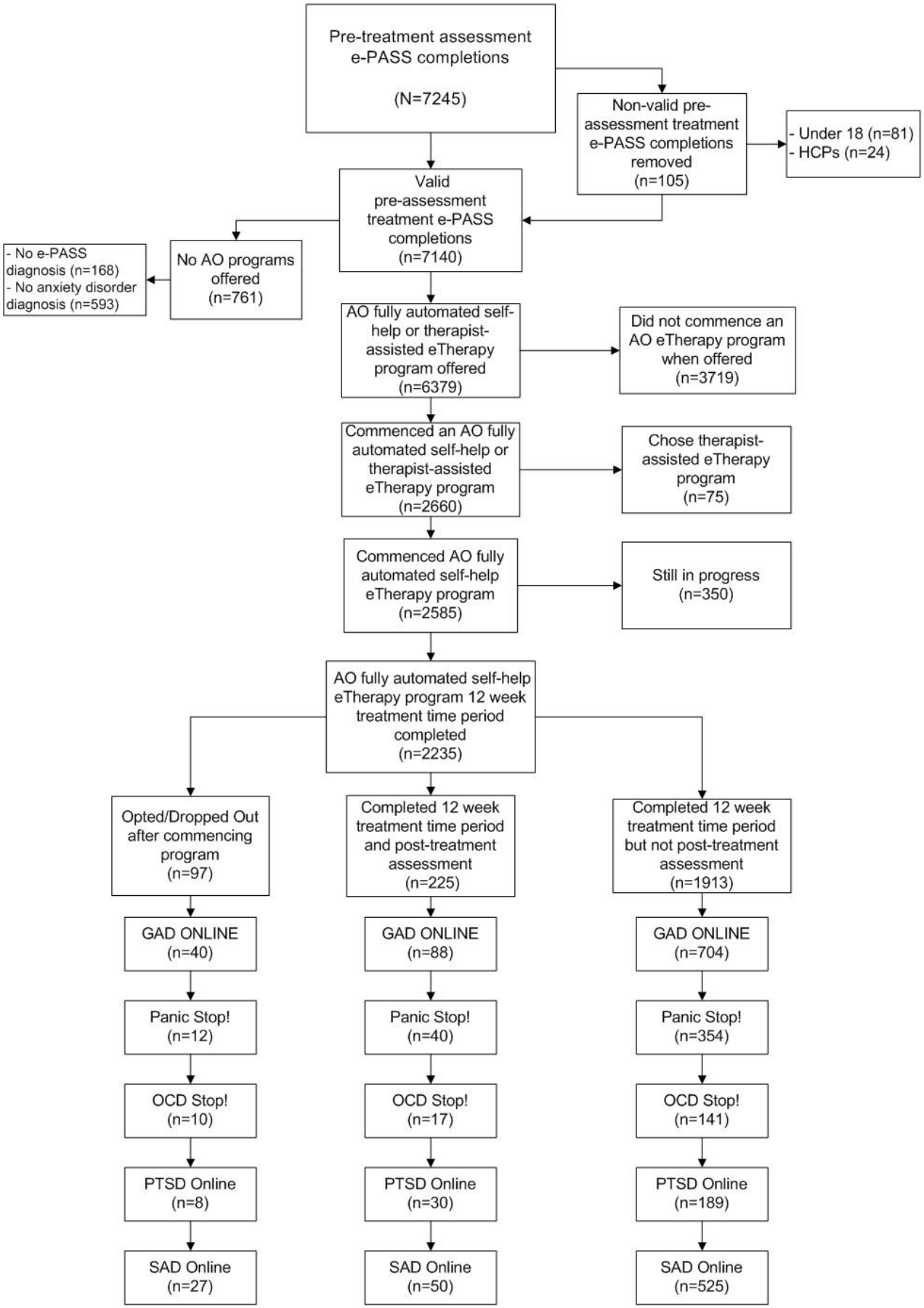
Attrition

The overall completion rate, for the purposes of this study, was defined as the number of participants who started a program and completed the 12-week posttreatment assessment (postassessment completion rate of 10.1% [225/2235], or 89.9% attrition rate). However, true attrition, as defined by those participants who opted out or dropped out of a program after commencement, was much lower at just 4% (97/2235). The bulk of participants (n = 1913) were those who commenced a program, did not opt or drop out, and did not complete their posttreatment assessment. In this situation the most reliable measure of attrition probably ignores this group, producing an attrition rate of 30.1% (225/322 = 69.9% completed). As Figure 1 shows, the attrition rates for the five treatment programs were fairly similar when this attrition measure was used (31% for GAD, 23% for Panic Stop!, 37% for OCD Stop!, 21% for PTSD and, 35% for SAD).

These data clearly illustrate the inherent completion difficulties facing e-mental health evaluation research using open-access research designs. It is therefore necessary to check for attrition bias using robust statistical techniques (eg, [45,46]). In cases where attrition bias is not found to be significant, the use of completer analysis is considered a legitimate and accurate means to analyze the data. However, in cases where significant attrition bias is found, the more conservative intention-to-treat method should be applied. Below we present two methods of checking for attrition bias, although either one is sufficient.

XSL•FO
RenderX

Figure 1. Recruitment flow (AO = Anxiety Online, e-PASS is an online psychological assessment system, GAD = generalized anxiety disorder, HCP = health care professional, OCD = obsessive–compulsive disorder, PTSD = posttraumatic stress disorder, SAD = social anxiety disorder).



XSL•FO
RenderX

Methods for Checking for Attrition Bias

We used two methods to check for attrition bias. The first method was initially proposed by Heckman [45]. In this approach, nominal logistic regression is used to predict the probability that each participant will complete the posttreatment assessment using a variety of pretreatment assessment and demographic measures. The Mills ratio is then produced using the ratio of the normal probability and cumulative distribution function for the residuals (1 minus the predicted probability for posttreatment assessment completion). This Mills ratio is then included as a covariate in a multivariate general linear model to determine the effect of the program on the posttreatment minus pretreatment changes in the outcome measures. If the Mills ratio is not significant, it indicates that the responses for those who did complete the posttreatment assessment are indicative of what could be expected for those who did not complete the posttreatment assessment; in other words, there is no attrition bias.

The second approach has been used by authors such as Rubin [46] to allow for propensity subclassification. In this study, propensity is the estimated probability of attrition developed using the nominal logistic regression procedure. In this study, we used the quartiles for this estimated probability to create subclassifications on which a multivariate general linear model is used to test for main effects and an interaction effect between the programs and the attrition propensity quartile effects. If no significant interaction effect and no significant quartile effect are found, it means that the program effects are similar across quartiles, suggesting that there is no significant attrition bias.

Check For Attrition Bias

We found nine pretreatment demographic variables to be significantly associated with attrition. As indicated in Table 1, both types of attrition (e-therapy program completed but no posttreatment assessment, and genuine attrition as defined by formally dropping our or opting out during treatment) were considered in this analysis. The results suggest that those who completed the posttreatment e-PASS tended to differ from the participants who did not in the following ways. It was more likely that on average the completers sought online assistance with the prime objective of finding a self-help program; were married or cohabiting with their partner; were not a homemaker, on a disability pension, or unemployed; were living in a regional area; were more likely to say that they had adequate support; were more likely to say that they learned by reading; had a lower pretreatment K6 score; had a higher age; and had fewer disorders diagnosed at pretreatment assessment. These differences made it necessary for special tests for attrition bias to be performed. These tests showed that none of the attrition-linked variables was associated with changes in the outcome variables, thereby confirming that there was no attrition bias.

XSL•FO
RenderX

Table 1. Predictor analysis for attrition categories

Variable	Attrition category						Test of association	
	No attrition: 12-week treatment period completed and posttreatment assessment (n = 225)		Opted out (n = 97)		12-week treatment period completed but not posttreatment assessment (n = 1913)		Test statistic value	*P* value
	%	n	%	n	%	n		
Reason for seeking online assistance							$\chi^2_2 = 14.6$	.001
To complete one of the self-help programs	66	149	52	50	53	1014		
Marital status							$\chi^2_{22} = 32.3$	.001
Married	43	96	40	39	35	668		
Single	24	55	27	26	29	545		
Cohabiting	17	39	13	13	19	367		
Other	16	35	20	19	17	331		
Employment status							$\chi^2_{12} = 31.6$	.002
Full-time	36	80	31	30	41	788		
Part-time	25	56	26	25	25	473		
Home, disability, or unemployed	19	46	24	23	22	411		
Retired	8	17	4	4	2	42		
Other	12	26	16	15	10	199		
Residential setting							$\chi^2_6 = 14.2$	.03
Metropolitan	62	139	77	75	67	1272		
Regional	27	61	17	16	22	411		
Rural	11	25	5	5	11	210		
Remote	0	0	1	1	1	17		
Adequate support							$\chi^2_2 = 8.2$	.02
Yes	50	113	33	32	46	878		
Preferred learning style							$\chi^2_2 = 12.3$	.06
Hearing	8	19	6	6	6	113		
Reading	36	82	31	30	29	545		
Looking	13	29	14	14	19	363		
Doing	42	95	49	47	47	891		
	Mean	SD	Mean	SD	Mean	SD		
Kessler-6	16.14	5.02	16.86	4.85	17.05	4.84		
Age (years)	42.08	12.51	37.96	13.12	36.64	11.94		
Number of disorders diagnosed	4.40	2.12	4.78	2.25	4.90	2.22		

We used a nominal logistic regression analysis to predict the attrition category for all participants on the basis of the above nine variables. The estimated probability of completion for the posttreatment assessment was saved for each person who actually completed the posttreatment assessment.

Using the Heckman [45] approach, the Mills ratio for the 225 people who completed the posttreatment assessment was calculated as described above, and a MANOVA was run for the change in all the metric outcome variables to test for differences in the program effects while controlling for the Mills ratio. We found that the Mills ratio had no significant effect ($F_{9,207} = .686$,

XSL•FO
RenderX

P = .72) and that there was no significant interaction between the program and the Mills ratio ($F_{36,840}$ = .854, P =.71). This confirms that there is unlikely to be any attrition bias for any of the e-therapy programs.

Next, we used the propensity subclassification approach [46] to split the sample of the 225 participants who completed the posttreatment assessment into four groups. The groups were differentiated in terms of the likelihood of attrition using the quartiles for this estimated probability and were of similar size. Effectively this differentiation controls for the likelihood of attrition, allowing us to determine whether there is attrition bias. A 2-way MANOVA was run for the change in all the outcome variables allowing for an interaction effect between the attrition propensity quartiles and the programs. Neither the interaction effect nor the main effect for the attrition propensity quartiles was found to be significant ($F_{108,1845}$ = 1.078, P = .28; $F_{27,597}$ = .810, P = .74), confirming that there is unlikely to be any attrition bias for any of the programs.

In summary, both techniques for assessing attrition bias delivered nonsignificant findings and demonstrate that attrition bias was highly unlikely for all five of the e-therapy programs. Given this result, we analyzed the data for each of the five fully automated e-therapy anxiety disorder treatment programs using a completer analysis.

Power Analysis

Target sample size required was determined by GPower [47]. To achieve power of 80% (alpha =.05), 34 participants per e-therapy treatment group were required to detect a moderate effect size on the primary outcome measure (clinical severity rating). Three of the five e-therapy treatment program groups had >34 participants (GAD Online, n = 88; SAD Online, n = 50; Panic Stop!, n = 40); however, PTSD Online and OCD Stop! reached only 30 and 17, respectively, so these results should be interpreted with greater caution.

Results

Participant Characteristics

A total of 225 people met the inclusion criteria and completed both pre- and posttreatment assessments. Across the five e-therapy programs, 69 men participated (69/225 = 31%) with the average age of all participants being 42.1 (SD 12.5) years (men, mean 44.5, SD 13.4; women, mean 41.0, SD 12.0 years). The overwhelming majority of participants were Australian residents (215/225 = 95.6%).

Table 2 presents demographic information for each of the five e-therapy program groups.

XSL•FO
RenderX

Table 2. Demographic results (n, %) of the posttreatment assessment completers by each one of the five Anxiety Online fully automated self-help e-therapy treatment programs

Demographic variable	GAD[a] Online (n = 88)		Panic Stop! (n = 40)		OCD[b] Stop! (n = 17)		PTSD[c] Online (n = 30)		SAD[d] Online (n = 50)	
	%	n	%	n	%	n	%	n	%	n
Gender										
Male	33	29	40	16	53	9	13	4	22	11
Female	67	59	60	24	47	8	87	26	78	39
Age category (years)										
18–24	8	7	5	2	6	1	10	3	10	5
25–34	18	16	20	8	24	4	23	7	34	17
35–44	27	24	40	16	41	7	20	6	20	10
45–54	23	20	20	8	12	2	30	9	20	10
55–64	21	18	13	5	18	3	13	4	10	5
65–74	2	2	3	1	0	0	3	1	4	2
75+	1	1	0	0	0	0	0	0	2	1
Marital status										
Single	19	17	18	7	24	4	37	11	32	16
Married	51	45	53	21	41	7	23	7	32	16
Cohabitating	17	15	18	7	24	4	10	3	20	10
In a relationship but not living together	7	6	10	4	0	0	3	1	8	4
Separated/divorced and not in a relationship	5	4	0	0	12	2	17	5	6	3
Widowed and not in a relationship	0	0	3	1	0	0	10	3	2	1
Other	1	1	0	0	0	0	0	0	0	0
Australian resident										
Yes	97	85	95	38	94	16	93	28	96	48
Residential setting										
Metropolitan	66	58	58	23	71	12	50	15	62	31
Regional	25	22	30	12	24	4	37	11	24	12
Rural	9	8	13	5	6	1	13	4	14	7
Remote	0	0	0	0	0	0	0	0	0	0
Secondary education										
Did not complete primary school	1	1	0	0	0	0	3	1	0	0
Completed primary school	0	0	0	0	0	0	0	0	0	0
Completed secondary up to year 9	1	1	3	1	0	0	7	2	4	2
Completed secondary year 10	10	9	15	6	0	0	13	4	14	7
Completed secondary year 11	5	4	5	2	0	0	10	3	8	4
Completed secondary year 12	83	73	78	31	100	17	67	20	74	37
Highest level of tertiary education										
None	8	7	25	10	6	1	17	5	12	6
Apprenticeship/trade	1	1	8	3	0	0	3	1	4	2
Other certificate	2	2	3	1	0	0	27	8	12	6
Diploma	11	10	3	1	6	1	17	5	8	4
Current undergraduate	11	10	10	4	0	0	3	1	10	5

XSL•FO
RenderX

Demographic variable	GAD[a] Online (n = 88)		Panic Stop! (n = 40)		OCD[b] Stop! (n = 17)		PTSD[c] Online (n = 30)		SAD[d] Online (n = 50)	
	%	n	%	n	%	n	%	n	%	n
Completed undergraduate	36	32	23	9	53	9	13	4	28	14
Postgraduate	28	25	28	11	35	6	17	5	16	8
Other	1	1	3	1	0	0	3	1	10	5
Employment status										
Employed full-time	36	32	40	16	47	8	30	9	30	15
Employed part-time/casual	26	23	35	14	29	5	23	7	14	7
Home duties	6	5	3	1	0	0	10	3	14	7
Disability support	0	0	3	1	12	2	3	1	2	1
Unemployed	7	6	8	3	6	1	17	5	20	10
Retired	9	8	8	3	6	1	3	1	8	4
Other	16	14	5	2	0	0	13	4	12	6
Currently taking an antidepressant or benzodiazepine medication?										
Yes	26	23	35	14	23	4	30	9	12	6
Currently receiving mental health assistance?										
Yes	42	37	50	20	41	7	57	17	22	11
Diagnosed physical health condition?										
Yes	40	35	33	13	35	6	47	14	36	18
Stage of change										
Not interested or no need at this time	0	0	3	1	0	0	0	0	0	0
Neither here nor there	0	0	3	1	0	0	10	3	6	3
Prepared to take action	57	50	43	17	59	10	53	16	48	24
Already making changes	39	34	33	13	35	6	30	9	34	17
Relapsed and looking for additional assistance	5	4	20	8	6	1	7	2	12	6
Do you feel you have an adequate level of social support or enragement in social/community activities?										
Yes	42	37	70	28	65	11	57	17	40	20
Preferred learning style										
Hearing	5	4	15	6	18	3	0	0	12	6
Reading	40	35	38	15	29	5	33	10	34	17
Looking/watching	10	9	15	6	12	2	13	4	16	8
Doing	46	40	33	13	41	7	53	16	38	19

[a] Generalized anxiety disorder.

[b] Obsessive–compulsive disorder.

[c] Posttraumatic stress disorder.

[d] Social anxiety disorder.

Treatment Outcomes

The number, means, standard deviations, *F* scores, *P* values, Cohen d, and confidence intervals for the five key dependent variables (per e-therapy program group) at the two assessment periods are shown in Table 3.

XSL•FO
RenderX

Table 3. Pre- and posttreatment assessment by Anxiety Online e-therapy treatment program group for posttreatment completers

Variable by e-therapy program disorder type	n	Mean	SD	$F_{1,n-1}$	*P* value	Cohen d (within groups)	95% CI[a]
GAD[b] Online	88						
GAD CDSR[c] pre[d]		3.26	1.5				
GAD CDSR post[e]		1.82	1.6	64.97	<.001	1.22	1.1 to 1.8
K6[f] pre		16.64	4.4				
K6 post		13.65	4.2	58.70	<.001	1.16	2.2 to 3.8
Disorder number[g] pre		4.24	1.8				
Disorder number post		3.17	1.9	34.45	<.001	0.89	0.7 to 1.4
Confidence[h] pre		3.12	0.9				
Confidence post		3.63	0.8	25.18	<.001	0.77	–0.7 to –0.3
Quality of life[i] pre		3.37	0.8				
Quality of life post		3.59	0.8	5.67	.02	0.36	–0.4 to –0.04
Panic Stop!	40						
PD[j] CDSR pre		3.13	1.9				
PD CDSR post		1.63	2.2	24.44	<.001	1.12	0.9 to 2.1
K6 pre		15.18	4.5				
K6 post		13.43	4.9	12.79	.001	0.81	0.8 to 2.7
Disorder number pre		4.60	2.2				
Disorder number post		3.97	2.6	3.92	.055	0.45	–0.01 to 1.3
Confidence pre		3.03	1.0				
Confidence post		3.48	0.9	11.32	.002	0.75	–0.7 to –0.2
Quality of life pre		3.55	1.0				
Quality of life post		3.60	1.0	0.50	.62	0.11	–0.3 to 0.2
OCD[k] Stop!	17						
OCD CDSR pre		2.33	0.9				
OCD CDSR post		1.52	1.8	4.95	.04	0.83	0.04 to 1.6
K6 pre		14.06	6.2				
K6 post		13.47	6.6	0.45	.51	0.23	–1.3 to 2.5
Disorder number pre		3.29	1.4				
Disorder number post		2.12	1.3	9.79	.006	1.08	0.4 to 2.0
Confidence pre		3.18	0.8				
Confidence post		3.76	0.9	11.59	.004	1.17	–1.0 to –0.2
Quality of life pre		3.71	1.1				
Quality of life post		4.00	1.1	6.67	.02	0.87	–0.5 to –0.1
PTSD[l] Online	30						
PTSD CDSR pre		3.17	1.6				
PTSD CDSR post		1.98	1.8	6.71	.02	0.72	0.3 to 2.1
K6 pre		18.53	5.2				
K6 post		14.20	5.7	13.54	.001	0.95	1.9 to 6.7
Disorder number pre		5.33	2.6				

XSL•FO
RenderX

Variable by e-therapy program disorder type	n	Mean	SD	$F_{1,n-1}$	*P* value	Cohen d (within groups)	95% CI[a]
Disorder number post		4.00	2.9	10.55	.003	0.85	0.5 to 2.2
Confidence pre		3.03	1.0				
Confidence post		3.83	0.9	18.08	<.001	1.08	–1.2 to –0.4
Quality of life pre		2.97	0.9				
Quality of life post		3.50	0.9	14.17	.001	0.96	–0.8 to –0.2
SAD[m] Online	50						
SAD CDSR pre		3.10	1.7				
SAD CDSR post		2.20	2.0	16.73	<.001	0.84	0.4 to 1.3
K6 pre		15.30	5.3				
K6 post		14.26	4.9	2.42	.13	0.31	–.3 to 2.4
Disorder number pre		4.32	2.3				
Disorder number post		3.74	2.3	6.33	.02	0.50	0.1 to 1.0
Confidence pre		2.90	1.0				
Confidence post		3.44	0.9	13.13	.001	0.70	–0.8 to –0.2
Quality of life pre		3.24	0.9				
Quality of life post		3.52	0.9	6.84	.01	0.51	–0.5 to –0.1

[a] Confidence interval (mean difference).

[b] Generalized anxiety disorder.

[c] e-PASS (online psychological assessment and referral system) clinical disorder severity rating, range 0–8.

[d] pre = preassessment.

[e] post = postassessment.

[f] Kessler6, range 6–30.

[g] Disorder number = number of disorders assessed by e-PASS, range 0–21.

[i] Quality-of-life ratings range 1–5.

[j] Panic disorder.

[k] Obsessive–compulsive disorder.

[l] Posttraumatic stress disorder.

[m] Social anxiety disorder.

GAD Online Program

For the GAD Online program a repeated measures MANOVA revealed a significant multivariate time effect for the e-PASS severity rating, K6, e-PASS total, quality-of-life, and confidence outcome measures ($F_{5,83} = 19.92$, $P < .001$). Follow-up repeated measures ANOVAs revealed significant improvements on all five variables. Three of the five treatment outcome variables produced large effect sizes, with one medium effect size and one small effect size (see Table 3).

Panic Stop! Program

For the Panic Stop! program a repeated measures MANOVA with these variables revealed a significant multivariate time effect ($F_{5,35} = 8.87$, $P < .001$). Follow-up repeated measures ANOVAs on three of the five variables revealed significant improvements on three variables. Two of these variables produced large effect sizes, with one medium effect size, one small effect size, and one very small (under .20) (see Table 3).

OCD Stop! Program

For the OCD Stop! program a repeated measures MANOVA with these variables revealed a significant multivariate time effect ($F_{5,12} = 4.21$, $P = .02$). Follow-up repeated measures ANOVAs revealed significant improvements on four of the five variables. Four of these variables produced large effect sizes with one small effect size (see Table 3).

PTSD Online Program

For the PTSD Online program a repeated measures MANOVA with these variables revealed a significant multivariate time effect ($F_{5,25} = 4.89$, $P = .003$). Follow-up repeated measures ANOVAs revealed significant improvements on all five variables. Four of these outcome variables produced large effect sizes with one medium effect size (see Table 3).

SAD Online Program

For the SAD Online program, a repeated measures MANOVA revealed a significant multivariate time effect for these variables ($F_{5,45} = 5.14$, $P = .001$). Follow-up repeated measures ANOVAs

XSL•FO
RenderX

revealed significant improvements on four of the five variables; however, only one of these variables produced a large effect size, with three medium effect sizes and one small effect size (see Table 3).

e-Therapy Program Treatment Satisfaction and Time Spent using the e-Therapy Program

Satisfaction with the e-therapy programs was rated as moderately high on average within all five groups (see Table 4), with the PTSD Online group obtaining the highest average score (3.73/5.00 = 74.6%). In terms of how much the participants liked their e-therapy program, average scores fell into the "somewhat" to "quite a lot" range, with the PTSD Online group likeability score the highest (5.67/8.00 = 70.9%).

Chi-square tests indicated no significant differences between the five e-therapy program groups on the two e-therapy program satisfaction questions (χ^2_{16} = 12.8, P = .69; χ^2_{16} = 16.3, P = .43). Posttreatment assessment completers were also asked how much time they spent using their respective e-therapy programs over the 12 weeks. GAD Online participants reported the most amount of time and the OCD Stop! participants the least amount of time (see Table 4). An ANOVA indicated no significant differences between the five e-therapy program groups with respect to the amount of time in minutes spent using their program over the 12-week treatment period ($F_{4,220}$ = 0.176, P = .95). The average total time in minutes across the five different e-therapy programs was 395.60 (SD 277.2) minutes or 6.59 hours over 12 weeks.

Table 4. e-Therapy program treatment satisfaction and likability ratings and time spent using their e-therapy program over the 12-week treatment period

Satisfaction and program usage variable	GAD[a] Online (n = 88)		Panic Stop! (n = 40)		OCD[b] Stop! (n = 17)		PTSD[c] Online (n = 30)		SAD[d] Online (n = 50)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
How satisfied are you with the online treatment program that you undertook?[e]	3.53	1.0	3.68	0.9	3.65	0.9	3.73	0.8	3.42	0.9
How much did you like the online program?[f]	5.23	2.3	5.45	1.8	5.41	1.8	5.67	1.6	5.20	1.8
How many hours did you spend in total reading/viewing the content in the online program? (minutes)	402.61	283.8	390.75	249.3	344.12	224.0	402.00	268.5	400.80	293.3

[a] Generalized anxiety disorder.

[b] Obsessive–compulsive disorder.

[c] Posttraumatic stress disorder.

[d] Social anxiety disorder.

[e] Rating scale: 1 = not at all, 2 = slightly, 3 = moderately, 4 = highly, 5 = very highly.

[f] Rating scale: 0 = not at all, 2 = a little, 4 = somewhat, 6 = quite a lot, 8 = very much so.

Reasons for Opting Out

Participants who opted out of the Anxiety Online program (n = 97) were asked to check one or more items regarding what barrier(s) prevented them from completing their Anxiety Online program when they opted out. Table 5 presents the item(s) endorsed.

XSL•FO
RenderX

Table 5. Endorsed barriers that prevented those who opted out from completing their Anxiety Online program (n = 97)

Barriers preventing program completion	%	n
None—got what I needed	29	28
Time pressures	21	20
Lack of motivation	14	14
Too anxious about the content	11	11
Became able to access face-to-face assistance	8	8
Realized I preferred face-to-face assistance	7	7
Internet connection or computer problems	5	5
The program did not seem very useful	5	5
Found the program boring	3	3
Found the program unhelpful	3	3
The program was too hard to navigate	3	3
The program material was too hard to understand	3	3
The program was going to take too long to do	2	2
There was too much text to read	2	2
Too anxious using the computer	1	1
The screen was hard to read (text was too small)	0	0

Basic Professional Labor Time Cost Analysis

The Australian Psychological Society [48] schedule of recommended fees for psychological services assumes the following costs: consultation session 45–60 minutes = AU $212; clinical assessment session 76–90 minutes = AU $308. If one were to use these fees to calculate the professional human labor time costs associated with all the e-PASS pretreatment assessments undertaken (n = 7140) and uptake of all the Anxiety Online e-therapy treatment programs since launch (n = 2563; 2660 – 97 opt-outs), the human labor time costing would equate to AU $8.7 million. The cost of developing Anxiety Online and ongoing maintenance over this period has been close to AU $2.0 million, with the bulk of this amount being a one-off development cost (AU $1.66 million). The labor time cost saving resulting from the Anxiety Online service in the first 18 months of operation is therefore estimated at AU $6.7 million, and into the future the cost savings should be greater, given that the start-up costs will not be recurring expenses.

Discussion

We observed significant reductions in the GAD, PD/A, OCD, PTSD, and SAD e-PASS diagnostic severity ratings specific to each e-therapy program group and increased self-confidence ratings in managing ones' own mental health care for all five e-therapy program groups. Cohen d within-group treatment effect sizes were in the high-medium to large categories. Quality-of-life ratings significantly increased for four of the five e-therapy program groupings, with Panic Stop! participants showing little improvement based on the mean scores. Total number of diagnoses was significantly reduced for four of the five programs, and the K6 scores significantly reduced for three of the five e-therapy programs. When looking at the treatment effect sizes for the five program groups over the five measures, 14 were large, six were medium, four were small, and one was under 0.20. These results compare very favorably with other self-help e-therapy treatment programs, which typically have effect sizes between 0.40 and 0.70 [15,31].

Overall, e-therapy treatment satisfaction ratings were good. The total average across all five e-therapy treatment program groups was 72% (3.60/5) for satisfaction and likeability was 67% (5.39/8). Interestingly, these percentages are comparable with those we see for our e-therapy programs provided with therapist assistance [8,14,16,49]. Taking treatment outcome and satisfaction results into account, PTSD Online and GAD Online appear to be the strongest performers of the five e-therapy programs. The total time spent using the fully automated e-therapy programs was under 7 hours over the 12-week treatment period (or just under 33 minutes per week on average).

For those participants who opted out, the main barriers endorsed were none (got what they needed; 29%), time pressures (21%), lack of motivation (14%), and feeling anxious about the content (11%). The least endorsed barriers related to computer anxiety (1%); too much text to read (2%); that the program would take too long to complete (2%); that the material was too hard to understand (3%); that the program was hard to navigate around (3%), unhelpful (3%), boring (3%), or perceived as being not useful (5%);, Internet or computer problems (5%), or preferring (7%) or being able to access to face-to-face therapy (8%). These figures generally support the idea that the Anxiety Online programs themselves are not harmful or detrimental to those participants who opted out. Rather, close to a third of those who opted out prematurely endorsed that they "got what they needed" before the 12-week treatment completion time, and just over a third endorsed time pressures and lack of motivation as reasons for noncompletion.

XSL•FO
RenderX

Although very crude, the basic cost comparison between the development and maintenance costs of Anxiety Online to date relative to the cost of professional human labor using traditional delivery models suggests a 77% saving of AU $6.7 million. However, Anxiety Online and other e-mental health platforms could easily sustain a 10-fold increase in usage without substantially affecting maintenance costs; therefore, the cost savings could be far higher than this into the future. When considered in light of other advantages of e-therapy such as increased consumer access to care due to the removal of traditional barriers such as time, cost, and geographic location restraints, the online-delivery model seems likely to play an increasingly prominent role in modern mental health systems.

Implications

The major implications of being able to effectively deliver psychological treatment via the Internet relate largely to accessibility. The availability of e-mental health services means that anyone in any location with Internet access can access treatment immediately and at the times and intensities they choose (ie, not limited to scheduled appointments). It also means greater access for those with no or limited access to treatment programs or mental health specialists (eg, living in rural areas, incarcerated), or those who move residences frequently (eg, itinerant workers, armed forces personnel) or have limited mobility (eg, with chronic physical illness, older, disabled). e-Mental health interventions may also help those with mental health problems who are reticent to present to services for reasons such as perceived stigma [50], although recent research supports that some stigmatizing attitudes actually lead to increased likelihood to seek professional help [51].

Limitations

This study has some limitations that should be taken into consideration. This study was conducted as an open-access, participant choice, naturalistic trial, so the lack of a control group makes it impossible to conclude whether the improvements are a result of the active e-therapy programs or merely a result of other effects. Nevertheless, this design limitation does represent “real-world” mental health services, as consumers are allowed to choose their treatment program rather than being randomly assigned to a treatment program in a randomized controlled trial. Power analyses indicted that the numbers required for the OCD Stop! and PTSD Online e-therapy programs were suboptimal and so some extra caution in interpretation needs to be taken. In previous studies we have used intention-to-treat analysis to address the issue of missing data. Intention-to-treat analysis is an overly conservative approach [52] that potentially underestimates the effectiveness of open-access e-therapy programs, whereas completer analysis may overestimate the effectiveness of these programs if there is attrition bias. Attrition bias was found to be nonsignificant for these data, so there are strong grounds for accepting that the completer analysis results obtained accurately reflect the true effectiveness of the Anxiety Online fully automated self-help programs.

Finally, four of the five dependent variables used in this study lack strong psychometric validation, and this needs to be taken into account when interpreting the results. Further replication studies are required that involve the use of several validated measures before we can definitively confirm that the Anxiety Online fully automated self-help e-therapy programs are effective. Future analyses with the individual Anxiety Online programs will be able to validate the present results with ideal standardized measures, as well as the aforementioned current formal e-PASS psychometric evaluation.

Future Directions

While results of this open trial are encouraging, it is important that they be followed by randomized controlled trials comparing all five e-therapy programs with both a waitlist control and current best-practice face-to-face treatment in order to unequivocally establish treatment effectiveness. We are running a randomized controlled trial of the OCD Stop! e-therapy program and will be subjecting the other four Anxiety Online e-therapy programs to the same rigorous testing. Furthermore, while e-therapy experts are developing guidelines for working with specific clinical populations (see [53]) as well as Internet intervention research guidelines [54], we also need more research to address the truly important questions about who this modality works best for because, like all forms of treatment delivery, it is unlikely to be universally appealing or effective. We are preparing another paper discussing our preliminary investigations regarding what variables predict attrition for the fully automated self-help e-therapy treatment programs. In addition, we will also be conducting qualitative studies that will include interviewing a random selection of those participants who commenced one of our treatment programs, did not opt out, and did not complete their postassessment. At this point it would be highly speculative to state why so many participants did not complete their postassessment after selecting one of the Anxiety Online programs, apart from pointing to the fully automated nature of the Anxiety Online system (ie, the complete absence of human-based screening, assessment, and therapeutic assistance) to prompt, encourage, and at times enforce adherence behaviors to the treatment protocol.

The Anxiety Online platform will soon be subsumed under the name Mental Health Online, given the addition of new e-therapy treatment programs for nonanxiety conditions (eg, depression, bulimia, insomnia, multidisorder) and several more to follow over the coming months and years (eg, problem gambling, drugs and alcohol, hoarding). We are also integrating other communication modalities into the therapist-assisted programs (ie, instant messaging, audio-only chat, video chat) and the use of 3-dimensional virtual reality platforms and collaborative work spaces. We will offer these new modes of communication and e-therapy training in 2012 and also plan to open up the Anxiety Online/Mental Health Online infrastructure to practitioners in Australia and, potentially, around the world. In the future it will also be possible to access the Anxiety Online/Mental Health Online programs through a national e-mental health portal instigated by the Australian Government Department of Health and Ageing [55]. This portal will bring together many of the evidence-based, yet fragmented, e-mental health interventions currently operating in Australia, thus making it far easier for mental health consumers to find and receive the most appropriate course of treatment with the associated level of assistance that is best for them.

XSL•FO
RenderX

Conclusions

The results of this open-access participant choice evaluation trial suggest that the Anxiety Online e-therapy programs are promising and effective treatments for people with subclinical and clinical diagnoses of GAD, PD/A, OCD, PTSD, and SAD. e-Mental health treatment-delivery formats are increasing accessibility to mental health care and appear to provide a highly cost-effective and sustainable treatment-delivery model. It is envisaged that e-mental health treatment programs will soon become a common feature of modern mental health systems, and that such a development will bring with it unprecedented levels of service provision to those in need of specialist mental health treatment.

Acknowledgments

The authors would like to acknowledge the contribution of and thank Pat Donelan, Andrew Smith, Belinda Williamson, Dr Jo Abbott, David Nguyen, Dr Joanna Mitchell, Sue Lauder, Dr Maja Nedeljkovic, Dr Richard Moulding, David Shields, Dr Lisa Ciechomski, Kerrie Shandley, Professor Gerhard Andersson, Professor Justin Kenardy, Dr Sonja March, Associate Professor Clare Rees, Associate Professor Nick Titov, and the members of the Anxiety Online Executive and Quality Assurance Committees. We would also like to acknowledge the memory of Professor Jeffrey Richards. We would also especially like to thank the Australian Government Department of Health and Ageing, who provided the funding for the development of the Anxiety Online service and its ongoing maintenance. e-PASS and the Anxiety Online e-therapy treatment programs are available to the international public at no cost (fully automated self-help). The therapist-assisted version is open only to Australian residents, and this is at a minimal cost. The funders of Anxiety Online (Australian Government Department of Health and Ageing) had no other involvement in this study or report.

Conflicts of Interest

None declared

Multimedia Appendix 1

Four Anxiety Online e-therapy program screenshots providing an example of the use of audio, animations, online activities, and video.

[PDF File (Adobe PDF File), 545KB - jmir_v13i4e89_app1.pdf]

References

1. Henderson S, Andrews G, Hall W. Australia's mental health: an overview of the general population survey. Aust N Z J Psychiatry 2000 Apr;34(2):197-205. [Medline: 10789524]
2. Australian Bureau of Statistics. National Survey of Mental Health and Wellbeing: Summary of Results. ABS Cat No. 4326.0. Canberra: ABS; 2008.
3. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602 [FREE Full text] [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
4. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
5. Burgess PM, Pirkis JE, Slade TN, Johnston AK, Meadows GN, Gunn JM. Service use for mental health problems: findings from the 2007 National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2009 Jul;43(7):615-623. [doi: 10.1080/00048670902970858] [Medline: 19530018]
6. Andrews G, Creamer M, Crino R, Hunt C, Lampe L, Page A. The Treatment of Anxiety Disorders: Clinician Guides and Patient Manuals. 2nd edition. Cambridge, UK: Cambridge University Press; 2003.
7. Parslow RA, Jorm AF. Who uses mental health services in Australia? An analysis of data from the National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2000 Dec;34(6):997-1008. [Medline: 11127632]
8. Klein B, Mitchell J, Gilson K, Shandley K, Austin D, Kiropoulos L, et al. A therapist-assisted Internet-based CBT intervention for posttraumatic stress disorder: preliminary results. Cogn Behav Ther 2009 Jun;38(2):121-131. [doi: 10.1080/16506070902803483] [Medline: 20183691]
9. Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K, Kavanagh D, et al. Oxford Guide to Low Intensity CBT Interventions (Oxford Guides in Cognitive Behavioural Therapy). Oxford, UK: Oxford University Press; 2010.
10. Christensen H. e-hub, Centre for Mental Health Research, Australian National University. Beacon Home Page URL: http://www.beacon.anu.edu.au/ [accessed 2011-05-27] [WebCite Cache ID 5yzKacLtO]
11. Christensen H, Murray K, Calear AL, Bennett K, Bennett A, Griffiths KM. Beacon: a web portal to high-quality mental health websites for use by health professionals and the public. Med J Aust 2010 Jun 7;192(11 Suppl):S40-S44. [Medline: 20528708]

XSL•FO
RenderX

12. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
13. Carlbring P, Nilsson-Ihrfelt E, Waara J, Kollenstam C, Buhrman M, Kaldo V, et al. Treatment of panic disorder: live therapy vs. self-help via the Internet. Behav Res Ther 2005 Oct;43(10):1321-1333. [doi: 10.1016/j.brat.2004.10.002] [Medline: 16086983]
14. Klein B, Austin D, Pier C, Kiropoulos L, Shandley K, Mitchell J, et al. Internet-based treatment for panic disorder: does frequency of therapist contact make a difference? Cogn Behav Ther 2009 Jun;38(2):100-113. [doi: 10.1080/16506070802561132] [Medline: 19306149]
15. Titov N, Andrews G, Choi I, Schwencke G, Mahoney A. Shyness 3: randomized controlled trial of guided versus unguided Internet-based CBT for social phobia. Aust N Z J Psychiatry 2008 Dec;42(12):1030-1040. [doi: 10.1080/00048670802512107] [Medline: 19016091]
16. Klein B, Mitchell J, Abbott J, Shandley K, Austin D, Gilson K, et al. A therapist-assisted cognitive behavior therapy internet intervention for posttraumatic stress disorder: pre-, post- and 3-month follow-up results from an open trial. J Anxiety Disord 2010 Aug;24(6):635-644. [doi: 10.1016/j.janxdis.2010.04.005] [Medline: 20447802]
17. Lange A, Rietdijk D, Hudcovicova M, van de Ven JP, Schrieken B, Emmelkamp PM. Interapy: a controlled randomized trial of the standardized treatment of posttraumatic stress through the internet. J Consult Clin Psychol 2003 Oct;71(5):901-909. [doi: 10.1037/0022-006X.71.5.901] [Medline: 14516238]
18. Christensen H, Griffiths KM, Mackinnon AJ, Kalia K, Batterham PJ, Kenardy J, et al. Protocol for a randomised controlled trial investigating the effectiveness of an online e health application for the prevention of Generalised Anxiety Disorder. BMC Psychiatry 2010;10:25 [FREE Full text] [doi: 10.1186/1471-244X-10-25] [Medline: 20302678]
19. Kenardy J, McCafferty K, Rosa V. Internet-delivered indicated prevention for anxiety disorders: a randomized controlled trial. Behav Cogn Psychother 2003;31:279-289. [doi: 10.1017/S1352465803003047]
20. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [doi: 10.1136/bmj.37945.566632.EE] [Medline: 14742346]
21. Proudfoot J, Ryden C, Everitt B, Shapiro DA, Goldberg D, Mann A, et al. Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul;185:46-54 [FREE Full text] [Medline: 15231555]
22. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
23. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187:456-461 [FREE Full text] [doi: 10.1192/bjp.187.5.456] [Medline: 16260822]
24. Ruwaard J, Schrieken B, Schrijver M, Broeksteeg J, Dekker J, Vermeulen H, et al. Standardized web-based cognitive behavioural therapy of mild to moderate depression: a randomized controlled trial with a long-term follow-up. Cogn Behav Ther 2009 Dec;38(4):206-221. [doi: 10.1080/16506070802408086] [Medline: 19221919]
25. Ritterband LM, Thorndike FP, Gonder-Frederick LA, Magee JC, Bailey ET, Saylor DK, et al. Efficacy of an Internet-based behavioral intervention for adults with insomnia. Arch Gen Psychiatry 2009 Jul;66(7):692-698 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.66] [Medline: 19581560]
26. Riper H, Kramer J, Smit F, Conijn B, Schippers G, Cuijpers P. Web-based self-help for problem drinkers: a pragmatic randomized trial. Addiction 2008 Feb;103(2):218-227. [doi: 10.1111/j.1360-0443.2007.02063.x] [Medline: 18199300]
27. Kiropoulos LA, Klein B, Austin DW, Gilson K, Pier C, Mitchell J, et al. Is internet-based CBT for panic disorder and agoraphobia as effective as face-to-face CBT? J Anxiety Disord 2008 Dec;22(8):1273-1284. [doi: 10.1016/j.janxdis.2008.01.008] [Medline: 18289829]
28. Griffiths KM, Farrer L, Christensen H. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med J Aust 2010 Jun 7;192(11 Suppl):S4-11. [Medline: 20528707]
29. White A, Kavanagh D, Stallman H, Klein B, Kay-Lambkin F, Proudfoot J, et al. Online alcohol interventions: a systematic review. J Med Internet Res 2010;12(5):e62 [FREE Full text] [doi: 10.2196/jmir.1479] [Medline: 21169175]
30. Andersson G, Ljótsson B, Weise C. Internet-delivered treatment to promote health. Curr Opin Psychiatry 2011 Mar;24(2):168-172. [doi: 10.1097/YCO.0b013e3283438028] [Medline: 21285706]
31. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008;26:109-160. [doi: 10.1080/15228830802094429]
32. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
33. e-hub e-mental Health Research & Development. Centre for Mental Health Research, Australian National University. e-hub Self-Help Programs for Mental Health & Wellbeing URL: http://www.ehub.anu.edu.au/welcome.php [accessed 2011-05-27] [WebCite Cache ID 5yzKEK4E1]

XSL•FO
RenderX

34. Centre for Emotional Health, Macquarie University. eCentreClinic URL: http://www.ecentreclinic.org/ [accessed 2011-08-21] [WebCite Cache ID 616CwDFrW]
35. National eTherapy Centre. Swinburne University of Technology. Anxiety Online Home Page URL: http://www.anxietyonline.org.au/ [accessed 2011-05-26] [WebCite Cache ID 5yypM2wx4]
36. Bennett K, Reynolds J, Christensen H, Griffiths KM. e-hub: an online self-help mental health service in the community. Med J Aust 2010 Jun 7;192(11 Suppl):S48-S52. [Medline: 20528710]
37. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR. 4th edition, text revision. Washington, DC: American Psychiatric Association; 2000.
38. Shandley K, Klein B, Kyrios M, Austin D, Ciechomski L, Murray G. Training postgraduate psychology students to deliver psychological services online. Aust Psychol 2011;46:120-125. [doi: 10.1111/j.1742-9544.2011.00034.x]
39. National eTherapy Centre. Swinburne University of Technology. Anxietyonline: Health Professional Enquiry URL: http://www.anxietyonline.org.au/health-professionals [accessed 2011-05-27] [WebCite Cache ID 5yzKHfAVy]
40. National eTherapy Centre. Swinburne University of Technology. Anxietyonline: Terms and Conditions URL: http://www.anxietyonline.org.au/terms-and-conditions [accessed 2011-05-27] [WebCite Cache ID 5yzKM1f2Z]
41. Nguyen D, Klein B, Meyer D, Austin D. The psychometric properties and user experience of an online clinical assessment schedule. 2011 Presented at: International Society for Research into Internet Interventions (ISRII) International Conference; April 6-8, 2011; Sydney, Australia.
42. Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, et al. Screening for serious mental illness in the general population. Arch Gen Psychiatry 2003 Feb;60(2):184-189 [FREE Full text] [Medline: 12578436]
43. Furukawa TA, Kessler RC, Slade T, Andrews G. The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. Psychol Med 2003 Feb;33(2):357-362. [Medline: 12622315]
44. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale, NJ: Erlbaum; 1988.
45. Heckman JJ. Sample selection bias as a specification error. Econometrica 1979;47(1):153-161.
46. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997 Oct 15;127(8 Pt 2):757-763. [Medline: 9382394]
47. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [Medline: 17695343]
48. Australian Psychological Society. APS. 2011. Fees and Rebates URL: http://www.psychology.org.au/community/fees_rebates/ [accessed 2011-05-27] [WebCite Cache ID 5yzLk6klT]
49. Klein B, Richards JC, Austin DW. Efficacy of internet therapy for panic disorder. J Behav Ther Exp Psychiatry 2006 Sep;37(3):213-238. [doi: 10.1016/j.jbtep.2005.07.001] [Medline: 16126161]
50. Corrigan P. How stigma interferes with mental health care. Am Psychol 2004 Oct;59(7):614-625. [doi: 10.1037/0003-066X.59.7.614] [Medline: 15491256]
51. Mojtabai R. Mental illness stigma and willingness to seek mental health care in the European Union. Soc Psychiatry Psychiatr Epidemiol 2010 Jul;45(7):705-712. [doi: 10.1007/s00127-009-0109-2] [Medline: 19680588]
52. Dasgupta A, Lawson KA, Wilson JP. Evaluating equivalence and noninferiority trials. Am J Health Syst Pharm 2010 Aug 15;67(16):1337-1343. [doi: 10.2146/ajhp090507] [Medline: 20689122]
53. Proudfoot J, Klein B, Andersson G, Carlbring P, Kyrios M, Munro C, et al. Guided CBT internet interventions: specific issues in supporting clients with depression, anxiety and co-morbid conditions. In: Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K, Kavanagh DJ, et al, editors. Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K, Kavanagh DJ, Klein B, Lau MA, Proudfoot J, editors. Oxford Guide to Low Intensity CBT Interventions. Oxford, UK: Oxford University Press; 2010:253-267.
54. Proudfoot J, Klein B, Barak A, Carlbring P, Cuijpers P, Lange A, et al. Establishing guidelines for executing and reporting internet intervention research. Cogn Behav Ther 2011;40(2):82-97. [doi: 10.1080/16506073.2011.573807]
55. Australian Government, Department of Health and Ageing. 2011 May 10. Health Budget 2011-2012 URL: http://www.health.gov.au/internet/budget/publishing.nsf/Content/budget2011-hmedia02.htm [accessed 2011-05-28] [WebCite Cache ID 5z18YXGnz]

Abbreviations

ANOVA: analysis of variance
CBT: cognitive behavior therapy
DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR. 4th edition, text revision
e-PASS: online psychological assessment system

XSL•FO
RenderX

Edited by G Eysenbach; submitted 26.08.11; peer-reviewed by N Titov, G Andrews; comments to author 13.09.11; revised version received 16.09.11; accepted 04.10.11; published 04.11.11.

Please cite as:
Klein B, Meyer D, Austin DW, Kyrios M
Anxiety Online—A Virtual Clinic: Preliminary Outcomes Following Completion of Five Fully Automated Treatment Programs for Anxiety Disorders and Symptoms
J Med Internet Res 2011;13(4):e89
URL: http://www.jmir.org/2011/4/e89/
doi:10.2196/jmir.1918
PMID:22057287

©Britt Klein, Denny Meyer, David William Austin, Michael Kyrios. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 04.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Changes in Depressive Symptoms, Social Support, and Loneliness Over 1 Year After a Minimum 3-Month Videoconference Program for Older Nursing Home Residents

Hsiu-Hsin Tsai[1], PhD; Yun-Fang Tsai[1,2], PhD

[1]School of Nursing, Chang Gung University, Tao-Yuan, Taiwan

[2]Department of Nursing, Chang Gung Memorial Hospital at Keelung, Keelung, Taiwan

Corresponding Author:
Yun-Fang Tsai, PhD
School of Nursing
Chang Gung University
259, Wen-Hwa 1st Road, Kwei-Shan
Tao-Yuan, 333
Taiwan
Phone: 886 3 2118800 ext 3204
Fax: 886 3 2118868
Email: yftsai@mail.cgu.edu.tw

Abstract

Background: A 3-month videoconference interaction program with family members has been shown to decrease depression and loneliness in nursing home residents. However, little is known about the long-term effects on residents' depressive symptoms, social support, and loneliness.

Objective: The purpose of this longitudinal quasi-experimental study was to evaluate the long-term effectiveness of a videoconference intervention in improving nursing home residents' social support, loneliness, and depressive status over 1 year.

Methods: We purposively sampled 16 nursing homes in various areas of Taiwan. Elderly residents (N = 90) of these nursing homes meeting our inclusion criteria were divided into an experimental (n = 40) and a comparison (n = 50) group. The experimental group received at least 5 minutes/week for 3 months of videoconference interaction with their family members in addition to usual family visits, and the comparison group received regular family visits only. Data were collected in face-to face interviews on social support, loneliness, and depressive status using the Social Support Behaviors Scale, University of California Los Angeles Loneliness Scale, and Geriatric Depression Scale, respectively, at four times (baseline, 3 months, 6 months, and 12 months after baseline). Data were analyzed using the generalized estimating equation approach.

Results: After the videoconferencing program, participants in the experimental group had significantly lower mean change in instrumental social support scores at 6 months (–0.42, P = .03) and 12 months (–0.41, P = .03), and higher mean change in emotional social support at 3 (0.74, P < .001) and 12 months (0.61, P = .02), and in appraisal support at 3 months (0.74, P = .001) after adjusting for confounding variables. Participants in the experimental group also had significantly lower mean loneliness and depressive status scores at 3 months (–5.40, P < .001; –2.64, P < .001, respectively), 6 months (–6.47, P < .001; –4.33, P < .001), and 12 months (–6.27, P = .001; –4.40, P < .001) compared with baseline than those in the comparison group.

Conclusion: Our videoconference program had a long-term effect in alleviating depressive symptoms and loneliness for elderly residents in nursing homes. This intervention also improved long-term emotional social support and short-term appraisal support, and decreased residents' instrumental social support. However, this intervention had no effect on informational social support.

(J Med Internet Res 2011;13(4):e93) doi:10.2196/jmir.1678

KEYWORDS

Videoconference; nursing home; elderly; social support; depression; loneliness

XSL•FO
RenderX

Introduction

Similar to other countries, Taiwan has more and more older people living in nursing homes. The number of nursing homes in Taiwan grew from 10 in 1995 to 372 in 2010 [1], indicating the great need for health care professionals trained in taking care of older institutionalized adults. Nursing home placement has been widely discussed in the literature as a stressful life event that challenges older people [2]. Older people who live in nursing homes have a higher prevalence of depression, which contributes to excessive morbidity [3], than do those who live in communities [4]. This prevalence of depression varied from 25% to 45% in Western countries [5,6] and was 52%–54% in Taiwan [7-9]. Many nursing home residents also experience loneliness [10], which has been associated with cognitive deterioration and hopelessness [11]. Thus, interventions have been used to enhance the quality of life in this group of people [12].

Among these interventions, social support is one of top importance because the social support systems older people use are closely related, both in quality and in quantity, to their health and quality of life [13,14]. Social support behavior falls into four categories: emotional, instrumental, informational, and appraisal support. Emotional support or affective assistance involves providing caring, empathy, love, and trust. Instrumental support refers to providing tangible goods and services or tangible aid, while informational support involves assistance with problem solving. Appraisal support or so-called affirmational support involves communicating information that is relevant to self-evaluation, rather than to problem solving [15,16].

Although most nursing home residents have become functionally dependent due to poor physical health, their psychosocial needs do not decrease [17]. In other words, social support is meaningful to them because it may provide emotional comfort. One important aspect of social support for older nursing home residents is continuous involvement of family members. However, one-third of nursing home residents were found to seldom have visitors [18]. If those older adults relocated to another nursing home, they had even fewer visitors [19]. However, support is not limited to in-person visits. Though family members may have limited time to visit residents in person, telephone calls can reduce residents' loneliness [20]. With rapid advances in telecommunication technology, real-time audiovisual linkups are now possible among multiple locations via affordable software and hardware [21]. Providing real-time audiovisual telecommunication systems to nursing home residents in Hong Kong [21] and Japan [22] has been shown to add a new dimension for the majority who lack the skills and capacities to adapt to the nursing home environment.

The benefits of videoconferencing in medicine have been recognized as a feasible way of delivering care to frail older people living with chronic diseases [20]. Videoconferencing has also been demonstrated as a feasible way to promote social interactions among nonspeaking people living in communities [20] and between the frail elderly and their caregivers [23]. Elderly nursing home residents in Taiwan were shown by Tsai and colleagues [24] to have significantly fewer depressive symptoms and less loneliness after a 3-month program of 5 minutes/week of videoconference interactions with family members. These studies demonstrate that videoconferencing is a feasible way for individuals living either in the community or in institutions to communicate. However, those studies had small sample sizes or used a cross-sectional design. Therefore, larger longitudinal studies are needed to make causal claims about the relationship between the effectiveness of videoconferencing and participants' depressive status and loneliness and to improve the generalizability of results.

To date, no empirical studies have used a longitudinal design to examine the effectiveness of videoconferencing for nursing home residents in Taiwan. Understanding the effectiveness of videoconferencing in Taiwanese nursing homes would fill the knowledge gap on this topic. Thus, the purpose of this study was to evaluate the long-term follow-up effectiveness of a videoconference intervention program on nursing home residents' social support, loneliness, and depressive status after a minimum 3-month videoconference program.

Methods

Design, Sample, and Setting

A quasi-experimental longitudinal design was used. Nursing homes from northern and central Taiwan were purposively selected based on three criteria: size (>65 beds), with Internet access, and accessible to the researchers. Among 23 medium-large nursing homes that met the recruitment criteria, 7 declined to participate in our study. The remaining 16 nursing homes (total beds =2190) were therefore randomly assigned to the comparison or experimental group.

The sample size was estimated based on the rule that 15 participants were needed for each variable [25]. Since we tested three major variables (depression, loneliness, and social support), the ideal sample size for this study was 45. Residents in the 16 nursing homes were recruited if they met the following criteria: (1) older than 60 years, (2) Mini-Mental State Examination (MMSE) score ≥16 for participants with no formal education or MMSE score >20 for those with at least a primary school education [26,27], and (3) wireless Internet access on their residence floor. Residents' family members had to have access to the Internet and be familiar with Internet communication programs such as Skype. These inclusion criteria were met by 423 residents, who were invited along with their family members to participate in the study. Family members of the majority of residents (n = 333, 78.7%) declined to participate, resulting in 50 participants in the comparison group and 40 in the experimental group. Residents in the experimental group used videoconferencing to communicate with their families plus their usual communication activities, whereas residents in the comparison group maintained their usual activities.

Videoconference Program

The videoconference program used laptops and Internet communication programs. Nursing home residents were asked to use the Internet at least once a week, with help from a trained

XSL•FO
RenderX

research assistant, who spent at least 5 minutes per week with each resident for the first 3 months during their scheduled videoconference visit. This weekly frequency was chosen to reflect the frequency of in-person visits to a nursing home resident for the majority of families [28,29], and 3 months was chosen to allow them time to adjust to the new videoconference program [30,31]. After 3 months, whenever residents wanted to have a video communication with their family, they were helped by the nursing home staff (nurses or nurses' aides) who were trained by the authors. The residents' main family contact person was a spouse, child, grandchild, or significant other. The communication programs used were Windows Live Messenger (MSN; Microsoft Corporation, Redmond, WA, USA) or Skype (Skype Technologies SA, Luxembourg) via a 2 M/256 K wireless modem run on a large-screen (15.6 inch) laptop.

Procedure

The study was approved by the Institutional Review Board of the authors' institution. After permission was granted from the nursing homes' administrators, details of our research procedure were posted at the entrance of each nursing home. This announcement indicated that residents or family members interested in participating in the study could directly contact the study personnel or nursing home staff. We also asked nursing home staff to talk with residents who met our study criteria and their family members about their willingness to participate in this study. Those who were interested in participation were contacted by the research assistant, who explained their rights, benefits, confidentiality, and responsibilities when participating in the study. After signing informed consent, residents and family members made appointments for videoconferences. Family members were phoned or emailed before the scheduled time to remind them of the appointment. The family of residents in the experimental group could continue their in-person or telephone visits as usual. Laptops were left in the nursing homes for 1 year. For the first 3 months, residents were helped by a trained research assistant to use the videoconference technology in a private room; for the next 9 months, residents were helped by trained nursing home staff. All residents in both the experimental and comparison groups completed questionnaires for demographic information (baseline only), depressive symptoms, loneliness, and social support at baseline and at 3, 6, 9, and 12 months.

Study Variables

Demographic indicators included participants' age, gender, marital status, educational background, duration of residency in the nursing home, and frequency of family visits. Residents' physical status and cognitive status were measured at baseline using the Barthel Index [32], which assesses performance of activities of daily living (ADLs), and MMSE [26], respectively. The MMSE cut-off score for severe cognitive deficit is ≥16 for participants without formal education and ³20 for those with at least a primary school education [27].

Depressive status was measured by the Geriatric Depression Scale (GDS) [33]. The GDS has 30 items with a yes/no response set. Possible scores range from 0 to 30. The GDS cut-off score for depressive symptom is >10 for mild depression and >20 for severe depression. The reliability of the GDS in a previous study of Taiwanese nursing home elderly was 0.91 [24] and in this study was 0.84.

Loneliness was measured by the revised University of California Los Angeles (UCLA) Loneliness Scale [34]. Responses to the 10 items on the UCLA Loneliness Scale are rated on a 4-point Likert scale from 1 (never) to 4 (always). Possible scores range from 20 to 80, with higher scores indicating higher perceived loneliness. The reliability of the UCLA Loneliness Scale was 0.87 in a study of institutionalized elderly in Taiwan [35] and was 0.92 in the current study.

Social support was measured by the Social Support Behaviors Scale with three subscales: social support network, quantity of social support, and satisfaction with social support [36]. Social support network was measured by the number of family members or friends who contacted residents and the number of contacts (either by phone or in person) during the previous week. The quantity of social support was measured by asking participants to rate each social support behavior (emotional, informational, instrumental, and appraisal support) offered by different providers (spouse, children, relatives, neighbors, and friends). Responses to this 14-item subscale are rated on a 5-point Likert scale, with higher scores indicating more support from each social resource. The subscale reliabilities for social support network, quantity of social support, and satisfaction with social support were 0.71, 0.92, and 0.77, respectively in Taiwanese nursing home residents [24], and 0.72, 0.89, and 0.79, respectively, in this study.

Family involvement with residents was confirmed by asking nursing home staff to record the number of visits and phone calls made to the residents. The duration of each videoconference interaction during the year was recorded by either the research assistant or nursing staff. Videoconference use was coded as the frequency of all videoconference interactions per month.

Data Analysis

All data were coded before being entered into a computer. Statistical analyses were performed using SPSS for Windows version 15.0 (IBM Corporation, Somers, NY, USA). Participants' demographic data were analyzed by descriptive statistics. Differences between groups were compared at four points (baseline, 3 months, 6 months, and 12 months) using multiple linear regression with the generalized estimating equation (GEE) method [37]. That is, we used the GEE method's multiple linear regression model (with the time and group interaction) to compare the time effects between two groups with or without adjusting for the effects of confounding variables such as resident's age and length of nursing home residency. Statistical significance was defined as $P < .05$.

Results

Participants' Characteristics

The 40 participants in the experimental group were on average 73.82 (SD 11.19) years old at baseline. The experimental group's use of videoconferencing decreased over time at 3, 6, and 12 months: mean (SD) 2.09 (1.46), 1.69 (1.37), and 1.14 (1.22), respectively. However, this decrease was not statistically

XSL•FO
RenderX

significant. The majority of participants were female (22/40, 55%) and widowed (29/40, 73%), and 35% (14/40) had graduated from primary school. Their average MMSE and Barthel Index scores at baseline were 23.51 (SD 3.93) and 65.68 (SD 22.55), respectively, indicating good cognitive status and above average performance of ADLs. They had an average of 3.69 (SD 2.09) children. About half of these participants (22/40, 55%) were visited by a family member at least once a week, and only 18% (7/40) seldom (less than once a month) had a family member visit them. Their average length of residency was 28.38 (SD 30.79) months (Table 1).

Table 1. Demographic characteristics of experimental and comparison groups

	Comparison group (n = 50)		Experimental group (n = 40)			*t*(*P* value)
Variable	n (%)	Mean (SD)	n (%)	Mean (SD)	χ^2(df[a]; *P* value)	df[a] = 88
Age (years)		79.26 (7.07)		73.82 (11.19)		10.78 (.01)
Gender					0.2 (1; .63)	
Male	20 (40)		18 (45)			
Female	30 (60)		22 (55)			
Marital status					6.6 (3; .16)	
Single	1 (2)		2 (5)			
Married	20 (40)		8 (20)			
Divorced	3 (6)		1 (2)			
Widowed	26 (52)		29 (73)			
Education					12.7 (4; .06)	
None/illiterate	29 (58)		9 (23)			
Primary	8 (16)		14 (35)			
Junior high school	2 (4)		2 (5)			
Senior high school	8 (16)		11 (28)			
≥ College	3 (6)		4 (10)			
Number of children		3.90 (2.05)		3.69 (2.09)		0.10 (.64)
Residency (months)		29.32 (28.58)		28.38 (30.79)		0.04 (.87)
Barthel Index		63.10 (23.62)		65.68 (22.55)		0.16 (.61)
MMSE		22.22 (3.93)		23.51 (3.93)		0.05 (.13)
In-person visits						
None/seldom	7 (14)		7 (18)			
Monthly	5 (10)		11 (28)			
Weekly (>2/month)	34 (68)		18 (45)			
Daily (>5/week)	4 (8)		4 (10)			
Telephone calls (number/week)						
0	36 (72)		24 (60)			
1	10 (20)		8 (20)			
2–6	3 (6)		6 (15)			
≥7	1 (2)		2 (5)			

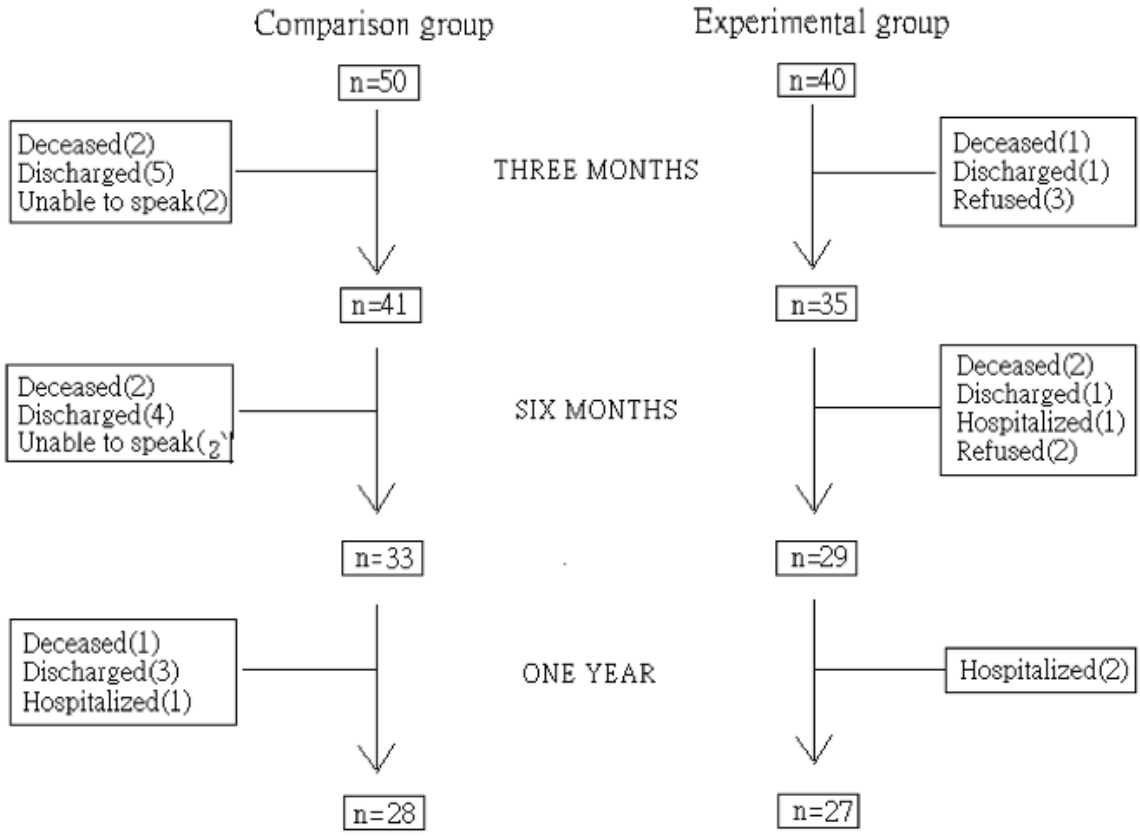
[a] Degrees of freedom.

The 50 participants in the comparison group were on average 79.26 (SD 7.07) years old at baseline. The majority were female (30/50, 60%), had no formal education (29/50, 58%), and were widowed (26/50, 52%). Their average MMSE and Barthel Index scores were 22.22 (SD 3.93) and 63.10 (SD 23.62), respectively, indicating good cognitive status and above average performance of ADLs. Their average number of children was 3.90 (SD 2.05). Most participants (38/50, 76%) were visited by a family member at least once a week and 14% (7/50) seldom had a family member visit them. Their average length of residency was 29.32 (SD 28.58) months (Table 1). The experimental and comparison group did not differ significantly in any demographic characteristics except for age (t = 10.78, P = .01).

XSL•FO
RenderX

During the 12-month study, 13 participants in the experimental group withdrew from the study (including 5 who declined to continue participating, 2 who relocated back home, and 3 who died), with an attrition rate of 33%. In the comparison group, 22 participants dropped out (including 5 who died, 12 who relocated back home, and 4 who developed cognitive deficits), with an attrition rate of 44% (Figure 1). Participants who dropped out and those who remained in the study did not differ significantly in any demographic characteristics.

Figure 1. Attrition of participants in the two groups over the 1-year study period.



Outcomes

Descriptive analysis of outcomes shows that both groups had the highest social support scores for the informational and instrumental social support subscales. The mean GDS scores for depressive status at baseline, and 3, 6, and 12 months were 12.75, 11.57, 12.85, and 13.00, respectively, for the experimental group, and 10.52, 10.56, 14.41, and 15.04, respectively, for the comparison group. These GDS scores did not differ significantly by independent *t* test between the two groups at any time (Table 2). Mean UCLA loneliness scores at baseline, and 3, 6, and 12 months were 49.70, 44.54, 46.21, and 45.92, respectively, for the experimental group, and 45.76, 45.59, 47.81, and 48.32, respectively, for the comparison group. The mean loneliness scores did not differ significantly between the two groups at any time (Table 2).

XSL•FO
RenderX

Table 2. Social support (Social Support Behaviors Scale), depressive status (Geriatric Depression Scale), and loneliness (University of California Los Angeles Loneliness Scale) scores by group at baseline, and 3, 6, and 12 months

Variable	Comparison group (n = 50)		Experimental group (n = 40)		*t* (*P* value) df[a] = 88
	Mean	SD	Mean	SD	
Social support: emotional support					
Baseline	9.76	1.69	9.49	1.57	0.78 (.59)
3 months	9.40	1.55	9.71	1.72	–0.80 (.66)
6 months	9.24	1.28	9.30	1.67	–0.14 (.24)
12 months	8.96	0.98	9.31	1.80	–2.30 (.04)
Social support: informational support					
Baseline	10.66	1.82	11.13	1.48	–1.31 (.21)
3 months	10.62	1.20	11.31	1.41	–2.33 (.02)
6 months	10.60	1.55	11.02	1.43	–1.08 (.61)
12 months	10.28	1.32	10.76	1.63	–1.16 (.25)
Social support: instrumental support					
Baseline	10.40	1.05	10.50	1.22	–0.38 (.33)
3 months	10.36	0.84	10.35	1.26	0.05 (.95)
6 months	10.53	1.05	10.31	1.14	0.77 (.51)
12 months	10.13	0.83	10.01	1.17	0.45 (.12)
Social support: appraisal support					
Baseline	9.25	1.60	9.03	1.25	0.73 (.14)
3 months	8.83	1.26	9.30	1.32	–1.59 (.12)
6 months	8.98	1.28	9.30	1.58	–0.87 (.39)
12 months	8.70	1.22	9.26	1.67	–1.42 (.16)
Total social support					
Baseline	141.54	18.64	141.85	17.26	–0.08 (.93)
3 months	139.00	14.95	143.11	18.36	–1.07 (.28)
6 months	139.55	16.08	140.43	18.02	–0.20 (.63)
12 months	134.08	12.74	138.08	19.36	–0.85 (.40)
Depressive status					
Baseline	10.52	4.06	12.75	5.50	1.94 (.06)
3 months	10.56	3.89	11.57	5.27	–0.96 (.34)
6 months	14.41	4.93	12.85	5.35	1.16 (.25)
12 months	15.04	4.61	13.00	4.50	1.60 (.11)
Loneliness					
Baseline	45.76	9.73	49.70	10.25	–1.85 (.09)
3 months	45.59	9.40	44.54	12.68	0.41 (.68)
6 months	47.81	9.97	46.21	11.87	0.57 (.56)
12 months	48.32	10.17	45.92	12.14	0.78 (.47)

[a] Degrees of freedom.

Time effects between the two groups were compared using the GEE method's multiple linear regression. As shown in Table 3, the average UCLA Loneliness Scale score of the experimental group was higher at baseline than that of the comparison group (beta = 3.94, $P = .09$). On the other hand, the changes in UCLA Loneliness Scale scores for the experimental group at 3, 6, and 12 months (compared with baseline) were significantly lower (–4.84, –6.46, and –6.42, respectively, all $P < .001$) than the corresponding changes in loneliness scores for the comparison group (–0.31, 2.81, and 2.77, respectively, with corresponding

XSL•FO
RenderX

P values = .55, .02, and .02) (Table 3). Moreover, after adjusting for the effects of residents' age and length of residency, all aforementioned results remained almost the same (right side of Table 3). Similarly, after controlling for residents' age and length of residency, the changes in GDS scores were on average significantly lower in the experimental group than in the comparison group at 3, 6, or 12 months (beta = –2.64, –4.33, and –4.40, respectively, all $P < .001$).

Table 3. Effects of videoconference intervention on participants' depressive status and loneliness at 3, 6, and 12 months in consideration of group, time, and time × group effects

Variable	Unadjusted				Adjusted[a]			
	beta	SE	χ^2_1	*P* value	beta	SE	χ^2_1	*P* value
Depressive status								
Group								
E vs C[b]	2.22	1.03	5.0	.06	2.26	1.00	5.1	.05
Time (vs B[c])								
3 months	0.97	0.33	8.9	<.001	0.99	0.32	9.3	<.001
6 months	6.95	0.63	123.4	<.001	6.99	0.63	142.7	<.001
12 months	7.64	0.70	118.9	<.001	7.71	0.70	119.6	<.001
Time × group[d]								
3 months	–1.36	0.56	6.0	.02	–2.64	0.57	21.3	<.001
6 months	–4.50	0.97	21.6	<.001	–4.33	1.03	17.6	<.001
12 months	–4.45	0.89	24.9	<.001	–4.40	0.92	23.1	<.001
Loneliness								
Group								
E vs C[b]	3.94	2.10	3.5	.09	3.27	2.23	2.2	.14
Time (vs B[c])								
3 months	–0.31	0.53	0.4	.55	–0.31	0.53	0.3	.56
6 months	2.81	1.23	5.2	.02	2.81	1.23	5.2	.02
12 months	2.77	1.22	5.2	.02	2.78	1.23	5.1	.02
Time × group[d]								
3 months	–4.84	1.14	18.0	<.001	–5.40	1.22	19.6	<.001
6 months	–6.46	1.64	15.4	<.001	–6.47	1.70	14.5	<.001
12 months	–6.42	1.64	15.3	<.001	–6.27	1.94	10.5	.001

[a] Adjusted for residents' age and length of residency.

[b] C: comparison group, E: experimental group.

[c] B: baseline measurement.

[d] Group 0: comparison group (reference group), group 1: experimental group.

This study had nine outcome variables of interest (depressive status, loneliness, total social support, emotional support, informational support, instrumental support, appraisal support, number of phone calls, and number of visits). Each outcome variable was analyzed exactly as in Table 3, but for each variable, we were mainly interested in comparing the time effects between the experimental and comparison groups. For simplicity, changes in only the time and group interaction terms are summarized in Table 4. The changes in appraisal and emotional social support scores at 3 months, after adjusting for the effects of residents' age and length of residency, were on average significantly higher in the experimental group than the corresponding changes in the comparison group (both beta = 0.74, *P*. = 001 and $P < .001$). The changes in instrumental social support scores at 6 and 12 months were on average significantly lower in the experimental group than in the comparison group (beta = –0.42 and –0.41, respectively, both $P = .03$) after adjusting for the effects of residents' age and length of residency (Table 4).

XSL•FO
RenderX

Table 4. Effects of videoconference intervention on participants' social support, depressive status, and loneliness at 3, 6, and 12 months in consideration of time × group effects

Variable	Unadjusted				Adjusted[a]			
	beta	SE	χ^2_1	*P* value	beta	SE	χ^2_1	*p* value
Total social support (time × group[b])								
3 months	3.71	2.10	3.1	.08	3.63	2.10	3.0	.09
6 months	–0.58	2.50	0.1	.82	–0.84	2.52	0.1	.74
12 months	–0.05	2.73	0.0	.99	–0.48	2.74	0.0	.86
Emotional support (time × group [b])								
3 months	0.60	0.19	10.1	.001	0.74	0.19	15.3	<.001
6 months	0.32	0.26	1.5	.23	0.40	0.28	2.1	.15
12 months	0.47	0.25	3.5	.06	0.61	0.26	5.3	.02
Informational support (time × group [b])								
3 months	–0.00	0.23	0.0	>.99	0.15	0.25	0.4	.53
6 months	–0.29	0.27	1.2	.28	–0.24	0.31	0.6	.44
12 months	–0.34	0.31	1.2	.28	–0.18	0.36	0.3	.62
Instrumental support (time × group [b])								
3 months	–0.20	0.14	2.0	.16	–0.14	0.15	0.9	.34
6 months	–0.47	0.18	6.7	.01	–0.42	0.19	4.7	.03
12 months	–0.41	0.19	4.9	.03	–0.41	0.19	4.6	.03
Appraisal support (time × group [b])								
3 months	0.66	0.22	9.3	.002	0.74	0.22	10.8	.001
6 months	0.37	0.24	2.3	.13	0.43	0.26	2.8	.10
12 months	0.57	0.31	3.5	.06	0.58	0.32	3.3	.07
Depressive status (time × group [b])								
3 months	–1.36	0.56	6.0	.02	–2.64	0.57	21.3	<.001
6 months	–4.50	0.97	21.6	<.001	–4.33	1.03	17.6	<.001
12 months	–4.45	0.89	24.9	<.001	–4.40	0.92	23.1	<.001
Loneliness (time × group [b])								
3 months	–4.84	1.14	18.0	<.001	–5.40	1.22	19.6	<.001
6 months	–6.46	1.64	15.4	<.001	–6.47	1.70	14.5	<.001
12 months	–6.42	1.64	15.3	<.001	–6.27	1.94	10.5	.001
Number of telephone calls (time × group [b])								
3 months	0.28	0.11	6.3	.01	0.28	0.12	5.8	.02
6 months	0.22	0.14	2.5	.11	0.20	0.15	1.7	.19
12 months	–0.01	0.15	0.0	.97	0.01	0.16	0.0	.95
Number of visits (time × group [b])								
3 months	–0.08	0.07	1.6	.21	–0.08	0.07	1.2	.27
6 months	0.05	0.08	0.4	.53	0.03	0.08	0.2	.70
12 months	0.05	0.08	0.4	.55	0.04	0.08	0.2	.67

[a] Adjusted for residents' age and length of residency.

[b] Group 0: comparison group (reference group), group 1: experimental group.

XSL•FO
RenderX

Discussion

This study demonstrated that our videoconference intervention alleviated elderly nursing home residents' perceived loneliness and improved their depressive status at 3, 6, and 12 months after the intervention. However, instrumental social support decreased at 6 and 12 months after the intervention.

Our 1-year attrition rate was high (35/90, 39%), as previously reported in similar longitudinal studies [38,39]. For example, an attrition rate of 39.5% was reported in a study on the effects of Internet use on health and depression [39]. Our attrition rate at 6 months (28/90, 31%) was also similar to the 6-month attrition rate (22.90%) reported for a study of elderly nursing home residents in Taiwan [38]. Among the reasons for case loss in our study, 16% (14/90) were discharged home, close to 15.26% as previously reported [38]. The majority of discharges were because the residents' families could not pay for the nursing home during the data collection period, which coincided with an economic depression in Taiwan. Other residents were discharged home because they were healthier than at admission.

Our research found that videoconferencing effectively improved elderly residents' depressive status at 3, 6, and 12 months. These results are consistent with a previous report [39] that using the Internet for communication with friends and family was associated with small but reliable decreases in depression. However, our study results are different from another report [40] of no significant difference in depression and loneliness among older adults after 5 months of training to access the Internet and email. In that study, however, participants were only trained to access the Internet, not to specifically contact family members or significant others [40]. The results of a previous study of institutionalized older Chinese adults [8] indicate that only family members can comfort these residents and reduce their depression and loneliness. This finding likely explains why depression and loneliness did not significantly change after intervention in White and colleagues' study [40]. In our program, not only were elderly residents shown how to use the Internet, but also appointments were arranged for them to communicate with their family members, who provide the majority of social support to Chinese elders and therefore reduce their depressive symptoms [8].

We found that videoconferencing effectively improved elderly residents' loneliness at 3, 6, and 12 months. These results are consistent with those of another study done in the United States [41] showing that loneliness was significantly reduced in 22 community-dwelling elderly people after 4 months of computer use. These results might be due to videoconference use providing a "social presence" to older adults [42]. For elderly residents in nursing homes, videoconferencing might add color to their lives. These results suggest that videoconference use is a good way to reduce loneliness of the elderly in both the community and institutions.

Our research found that videoconferencing, a computer-mediated communication, had no effects on instrumental social support at the 3-month data collection time, as previously reported [43,44]. However, we found that instrumental social support decreased significantly over time, but not the number of family members' in-person visits. In other words, family members provided less instrumental social support in terms of specific items and assistance, even though they kept visiting the elderly residents in person. After a long stay in a nursing home, elderly residents tend to adapt to the environment and not need extra items for daily life, since such things are provided by the nursing home. Thus, family members might not see instrumental social support as the best way to show their filial piety, or the elders might not ask family members to bring them things. These possibilities need to be examined in future studies.

Our videoconference intervention also improved emotional social support at 3 and 12 months and appraisal support after 3 months. The effect of videoconferencing on appraisal and emotional support at 3 months is similar to our previous study [24]. The lack of intervention effects on appraisal social support at 6 and 12 months might be due to decreased novelty and quality of videoconferencing. Although videoconferencing is a convenient way to connect with people at a distance, videoconferencing alone cannot improve the quality and amount of communications between people. In particular, when communicators view the communication as an obligation, they might feel bored, shorten the communication, or show an unpleasant attitude or tone.

From this point of view, we suggest that nursing home administrators increase the quality of communication by developing an interaction program such as arranging for family members to have a meal with residents at the nursing home and have a meal together via videoconference. One explanation for the long-term (12 months) decrease in emotional social support might be that nursing home residents feel safe or comforted by using videoconferencing as an alternative "social presence" so that they can immediately see their family member, even at a distance. Videoconferencing may offer them a chance to be part of family life. They also might feel comforted by seeing their family member's actual state and would not worry that the family member was hiding a problem to allay anxiety if he or she could not visit [45].

The use of videoconference visits decreased over time. This decreased use of videoconferencing was likely due to a loss in the novelty of videoconferencing, lack of staff to help the residents operate the devices, and a need to remind family members to use videoconferencing (busy family members tended to forget to use videoconferencing). However, these possible reasons for decreased videoconference use need to be supported by further research. Furthermore, we found that videoconference use was high in some nursing homes, especially for those residents with relatives living overseas. However, our data were not significant due to the small sample size from each nursing home. We found no studies on the relationships between videoconference use and the characteristics of nursing home residents' families. We therefore suggest that future research explore the relationships between videoconference use and characteristics of nursing home residents' family members, factors influencing videoconference use, effects of videoconferencing on the health of elderly residents and their families, and the cost effectiveness of the videoconference program.

XSL•FO
RenderX

Although the experimental and control groups did not differ significantly in any dependent variable at any time point, the experimental group showed significant changes in depression, loneliness, and two social support measures over time compared with the control group after controlling for residents' age and length of residency. The independent *t* test used to compare the results for each outcome variable at each measurement time point (Table 2) did not benefit from the strength of repeated measurements within participants. One possible reason that the independent *t* test did not show significant differences in mean outcome measurements between groups is that it did not include other repeated measurements from the same participants. Another possible reason is that the sample was too small to detect significant differences. On the other hand, the results of GEE analysis showed significant incremental changes in the dependent variables over time due to the likely impact of within-participant's repeated measurements.

Our research also showed that, after adjustment for residents' age and length of residency, the time effects between the experimental and comparison groups remained the same for all outcome variables except emotional social support at 12 months. In other words, after adjustment for time and group effects, age and length of residency had almost no significant impact on all outcome variables, except emotional social support at 12 months.

Furthermore, the outcome variables of loneliness, lack of social support, and depression might have been associated with each other. However, when we analyzed each variable by GEE method with and without controlling for other variables, we found that the trends were not affected (data not shown). Further research is recommended to explore the associations among these variables and their possible impact on the time effects.

Conclusion

Acknowledgments

Our videoconferencing program was funded by the National Science Council, an agency of the Taiwan government (NSC97-2314-B-182-018), to Hsiu-Hsin Tsai, Principle Investigator. We thank Professor Hsiu-Hung Wang, Professor Hao-Hua Chu, and Associate Professor Hsu-Min Tseng for their contributions to the program development. We also thank Professor Yue-Cune Chang for statistical consultation.

Conflicts of Interest

None declared

References

1. Taiwan Association of Long-Term Care Professionals. 2010. Institution of long-term care in Taiwan URL: http://www.ltcpa.org.tw/org/org.php [accessed 2011-10-05] [WebCite Cache ID 62CyNNjV3]
2. Haight BK, Michel Y, Hendrix S. Life review: preventing despair in newly relocated nursing home residents short- and long-term effects. Int J Aging Hum Dev 1998;47(2):119-142. [Medline: 9836092]
3. Parmelee PA, Katz IR, Lawton MP. Incidence of depression in long-term care settings. J Gerontol 1992 Nov;47(6):M189-M196. [Medline: 1430853]
4. Gueldner SH, Loeb S, Morris D, Penrod J, Bramlett M, Johnston L, et al. A comparison of life satisfaction and mood in nursing home residents and community-dwelling elders. Arch Psychiatr Nurs 2001 Oct;15(5):232-240. [doi: 10.1053/apnu.2001.27020] [Medline: 11584352]
5. McCurren C, Dowe D, Rattle D, Looney S. Depression among nursing home elders: testing an intervention strategy. Appl Nurs Res 1999 Nov;12(4):185-195. [Medline: 10589107]
6. Brodaty H, Draper B, Saab D, Low LF, Richards V, Paton H, et al. Psychosis, depression and behavioural disturbances in Sydney nursing home residents: prevalence and predictors. Int J Geriatr Psychiatry 2001 May;16(5):504-512 [FREE Full text] [Medline: 11376467]
7. Lin LC, Wang TG, Chen MY, Wu SC, Portwood MJ. Depressive symptoms in long-term care residents in Taiwan. J Adv Nurs 2005 Jul;51(1):30-37 [FREE Full text] [doi: 10.1111/j.1365-2648.2005.03457.x] [Medline: 15941458]
8. Tsai YF, Chung JW, Wong TK, Huang CM. Comparison of the prevalence and risk factors for depressive symptoms among elderly nursing home residents in Taiwan and Hong Kong. Int J Geriatr Psychiatry 2005 Apr;20(4):315-321. [doi: 10.1002/gps.1281] [Medline: 15799085]
9. Tsai YF, Wong TK, Tsai HH, Ku YC. Self-worth therapy for depressive symptoms in older nursing home residents. J Adv Nurs 2008 Dec;64(5):488-494. [doi: 10.1111/j.1365-2648.2008.04804.x] [Medline: 19146517]
10. Yeh SH, Lin LW, Lo SK. A longitudinal evaluation of nursing home care quality in Taiwan. J Nurs Care Qual 2003;18(3):209-216. [Medline: 12856905]
11. Hicks TJ. What is your life like now? Loneliness and elderly individuals residing in nursing homes. J Gerontol Nurs 2000 Aug;26(8):15-19. [Medline: 11276607]
12. York JL, Calsyn RJ. Family involvement in nursing homes. Gerontologist 1977 Dec;17(6):500-505. [Medline: 924171]

XSL•FO
RenderX

13. Shin JK, Kim KW, Park JH, Lee JJ, Huh Y, Lee SB, et al. Impacts of poor social support on general health status in community-dwelling korean elderly: the results from the korean longitudinal study on health and aging. Psychiatry Investig 2008 Sep;5(3):155-162. [doi: 10.4306/pi.2008.5.3.155] [Medline: 20046359]
14. Wang HH, Wu SZ, Liu YY. Association between social support and health outcomes: a meta-analysis. Kaohsiung J Med Sci 2003 Jul;19(7):345-351. [doi: 10.1016/S1607-551X(09)70436-X] [Medline: 12926520]
15. Barrera MJ. Social support in the adjustment of pregnant adolescents: assessment issues. In: Gottlieb BH, editor. Social Networks and Social Support. Beverly Hills, CA: Sage; 1981:69-96.
16. Langford CP, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. J Adv Nurs 1997 Jan;25(1):95-100. [Medline: 9004016]
17. Williams K, Kemper S, Hummert ML. Improving nursing home communication: an intervention to reduce elderspeak. Gerontologist 2003 Apr;43(2):242-247. [Medline: 12677081]
18. Gueldner SH, Clayton GM, Schroeder MA, Butler S, Ray J. Environmental interaction patterns among institutionalized and non-institutionalized older adults. Phys Occup Ther Geriatr 1993;11(1):37-53.
19. Barry JT, Miller DB. The nursing home visitor: who, when, where and for how long? Long Term Care Health Serv Adm Q 1980;4(4):261-274. [Medline: 10249087]
20. Hine N, Arnott JL. A multimedia social interaction service for inclusive community living: initial user trials. Univ Access Inf Soc 2002;2:8-17. [doi: 10.1007/s10209-002-0032-8]
21. Hui E, Woo J, Hjelm M, Zhang YT, Tsui HT. Telemedicine: a pilot study in nursing home residents. Gerontology 2001;47(2):82-87. [Medline: 11287732]
22. Nakamura K, Takano T, Akao C. The effectiveness of videophones in home healthcare for the elderly. Med Care 1999 Feb;37(2):117-125. [Medline: 10024116]
23. Savolainen L, Hanson E, Magnusson L, Gustavsson T. An Internet-based videoconferencing system for supporting frail elderly people and their carers. J Telemed Telecare 2008;14(2):79-82. [doi: 10.1258/jtt.2007.070601] [Medline: 18348753]
24. Tsai HH, Tsai YF, Wang HH, Chang YC, Chu HH. Videoconference program enhances social support, loneliness, and depressive status of elderly nursing home residents. Aging Ment Health 2010 Nov;14(8):947-954. [doi: 10.1080/13607863.2010.501057] [Medline: 21069600]
25. Mertens DM. Research Methods in Education and Psychology: Integrating Diversity with Quantitative and Qualitative Approaches. Thousand Oaks, CA: Sage; 1998.
26. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [Medline: 1202204]
27. Liu GK, Tai DD, Lin RT, Lai CL. Epidemiology of dementia in Taiwan [in Mandarin]. Res Appl Psychol 2000;7:157-169.
28. Bitzan JE, Kruzich JM. Interpersonal relationships of nursing home residents. Gerontologist 1990 Jun;30(3):385-390. [Medline: 2354798]
29. Port CL, Gruber-Baldini AL, Burton L, Baumgarten M, Hebel JR, Zimmerman SI, et al. Resident contact with family and friends following nursing home admission. Gerontologist 2001 Oct;41(5):589-596. [Medline: 11574703]
30. Brooke V. How elders adjust. Geriatr Nurs 1989;10(2):66-68. [Medline: 2703148]
31. Mikhail ML. Psychological responses to relocation to a nursing home. J Gerontol Nurs 1992 Mar;18(3):35-39. [Medline: 1556396]
32. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J 1965 Feb;14:61-65. [Medline: 14258950]
33. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37-49. [Medline: 7183759]
34. Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J Pers Soc Psychol 1980 Sep;39(3):472-480. [Medline: 7431205]
35. Tsai YF. Self-care management and risk factors for depressive symptoms among Taiwanese institutionalized older persons. Nurs Res 2007;56(2):124-131. [doi: 10.1097/01.NNR.0000263973.31748.0c] [Medline: 17356443]
36. Hsiung HF. Health Status, Social Support and Related Factors Among Community-Dwelling Elderly Living Alone [master's thesis, in Mandarin]. Tapei, Taiwan: National Taiwan University; 1999.
37. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
38. Chang SH, Fang MC, Chang HY. Enhancing three bliss concepts among nursing home elders in Taiwan. J Clin Nurs 2010 Mar;19(5-6):682-690. [doi: 10.1111/j.1365-2702.2009.03076.x] [Medline: 20500310]
39. Bessière K, Pressman S, Kiesler S, Kraut R. Effects of internet use on health and depression: a longitudinal study. J Med Internet Res 2010;12(1):e6 [FREE Full text] [doi: 10.2196/jmir.1149] [Medline: 20228047]
40. White H, McConnell E, Clipp E, Branch LG, Sloane R, Pieper C, et al. A randomized controlled trial of the psychosocial impact of providing internet training and access to older adults. Aging Ment Health 2002 Aug;6(3):213-221. [doi: 10.1080/13607860220142422] [Medline: 12217089]
41. Shapira N, Barak A, Gal I. Promoting older adults' well-being through Internet training and use. Aging Ment Health 2007 Sep;11(5):477-484. [doi: 10.1080/13607860601086546] [Medline: 17882585]

XSL•FO
RenderX

42. Cukor P, Baer L, Willis BS, Leahy L, O'Laughlen J, Murphy M, et al. Use of videophones and low-cost standard telephone lines to provide a social presence in telepsychiatry. Telemed J 1998;4(4):313-321. [Medline: 10220471]
43. Chang HJ, Ho MH, Huang CJ. Typology of online social support: a study of bulletin board system in Taiwan [in Mandarin]. Mass Commun Res 2008;94:61-105.
44. Walther JB, Park MR. Cues filtered out, cues filtered in: computer-mediated communication and relationships. In: Knapp ML, Daly JA, editors. Handbook of Interpersonal Communication, 3rd edition. Thousand Oaks, CA: Sage; 2002:529-563.
45. Tsai HH, Tsai YF. Older nursing home residents' experiences with videoconferencing to communicate with family members. J Clin Nurs 2010 Jun;19(11-12):1538-1543. [doi: 10.1111/j.1365-2702.2010.03198.x] [Medline: 20384665]

Edited by G Eysenbach; submitted 25.10.10; peer-reviewed by S Kossman; comments to author 09.05.11; revised version received 09.06.11; accepted 06.07.11; published 15.11.11.

Please cite as:
Tsai HH, Tsai YF
Changes in Depressive Symptoms, Social Support, and Loneliness Over 1 Year After a Minimum 3-Month Videoconference Program for Older Nursing Home Residents
J Med Internet Res 2011;13(4):e93
URL: http://www.jmir.org/2011/4/e93/
doi:10.2196/jmir.1678
PMID:22086660

©Hsiu-Hsin Tsai, Yun-Fang Tsai. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of Length or Relevance of Questionnaires on Attrition in Online Trials: Randomized Controlled Trial

Jim McCambridge[1]; Eleftheria Kalaitzaki[2]; Ian R White[3]; Zarnie Khadjesari[4]; Elizabeth Murray[4]; Stuart Linke[4]; Simon G Thompson[3]; Christine Godfrey[5]; Paul Wallace[4]

[1]Faculty of Public Health & Policy, London School of Hygiene & Tropical Medicine, London, United Kingdom

[2]MRC General Practice Research Framework, London, United Kingdom

[3]MRC Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom

[4]E-health Unit, Research Department of Primary Care and Population Health, University College London, London, United Kingdom

[5]Department of Health Sciences and HYMS, University of York, York, United Kingdom

Corresponding Author:
Jim McCambridge
Faculty of Public Health & Policy
London School of Hygiene & Tropical Medicine
15 - 17 Tavistock Place
London, WC1H 9SH
United Kingdom
Phone: 44 (0)20 7927 2945
Fax: 44 (0)20 7927 7958
Email: Jim.McCambridge@lshtm.ac.uk

Abstract

Background: There has been limited study of factors influencing response rates and attrition in online research. Online experiments were nested within the pilot (study 1, n = 3780) and main trial (study 2, n = 2667) phases of an evaluation of a Web-based intervention for hazardous drinkers: the Down Your Drink randomized controlled trial (DYD-RCT).

Objectives: The objective was to determine whether differences in the length and relevance of questionnaires can impact upon loss to follow-up in online trials.

Methods: A randomized controlled trial design was used. All participants who consented to enter DYD-RCT and completed the primary outcome questionnaires were randomized to complete one of four secondary outcome questionnaires at baseline and at follow-up. These questionnaires varied in length (additional 23 or 34 versus 10 items) and relevance (alcohol problems versus mental health). The outcome measure was the proportion of participants who completed follow-up at each of two follow-up intervals: study 1 after 1 and 3 months and study 2 after 3 and 12 months.

Results: At all four follow-up intervals there were no significant effects of additional questionnaire length on follow-up. Randomization to the less relevant questionnaire resulted in significantly lower rates of follow-up in two of the four assessments made (absolute difference of 4%, 95% confidence interval [CI] 0%-8%, in both study 1 after 1 month and in study 2 after 12 months). A post hoc pooled analysis across all four follow-up intervals found this effect of marginal statistical significance (unadjusted difference, 3%, range 1%-5%, $P = .01$; difference adjusted for prespecified covariates, 3%, range 0%-5%, $P = .05$).

Conclusions: Apparently minor differences in study design decisions may have a measurable impact on attrition in trials. Further investigation is warranted of the impact of the relevance of outcome measures on follow-up rates and, more broadly, of the consequences of what we ask participants to do when we invite them to take part in research studies.

Trial registration: ISRCTN Register 31070347; http://www.controlled-trials.com/ISRCTN31070347/31070347 Archived by WebCite at (http://www.webcitation.org/62cpeyYaY)

(J Med Internet Res 2011;13(4):e96) doi:10.2196/jmir.1733

KEYWORDS

Attrition; retention; missing data; response rates; alcohol; online

XSL•FO
RenderX

Introduction

A large multidisciplinary experimental literature has developed over many decades in which a wide range of methods to increase response rates in postal surveys have been evaluated [1]. Edwards and colleagues included 481 trials in their updated systematic review of this literature, which includes both postal and electronic surveys [2]. Among the methods identified to be effective in postal surveys are using shorter questionnaires (pooled odds ratios [ORs] for 56 trials for responding to shorter vs longer questionnaires = 1.64, 95% confidence interval [CI] 1.43-1.87) and asking more relevant questions (pooled ORs for 3 trials = 2.00, 95% CI 1.32-3.04). The findings from 32 randomized controlled trials of electronic surveys have been broadly similar to those obtained for postal surveys [2].

It is unclear to what extent findings on methods effective in enhancing response rates in surveys can be applied to studies involving follow-up. Attrition prevention may involve issues that are different from those concerned with maximizing survey response rates because being interviewed or providing questionnaire data some time after study entry is likely to be influenced by the history of study involvement and the demands it makes upon the participant.

There is not, however, a clear evidence base on effective methods to prevent loss to follow-up specifically in the contexts of cohort studies and trials. A 2007 systematic review of retention strategies in health care research found no studies that "explicitly compared the effectiveness of different retention strategies" [3]. A 2009 systematic review of drug trials for weight loss found that the number of attendances for research purposes made no difference to attrition [4]. In the same year, a meta-analysis found a range of study design characteristics to influence attrition in trials of antidepressants among older people [5].

It is important to minimize attrition in these types of studies, as participants lost to follow-up may have characteristics different from participants retained by the study, thus potentially introducing bias. Attrition is particularly problematic for online trials as it is usually substantial [6-7] and may differ between randomized groups, thus engendering additional difficulties in interpretation of study findings [8]. Attrition in online trials is also well known to be a more complex phenomenon than in conventional trials [6-7]. The online context permits intervention nonusage, which can be very high. This is often closely related to loss to follow-up for research purposes. It is this latter form of attrition that is the subject of the present study.

In planning the Down your Drink randomized controlled trial (DYD-RCT), as an attrition reduction measure, we decided to reduce the assessment burden by randomly allocating participants to complete only one of four secondary outcome questionnaires rather than all four [9]. Although this decision sacrificed statistical power to detect effects on these particular outcomes, it added to the capacity to detect effects on the primary outcome to the extent that reducing the overall assessment burden enhances follow-up rates [6]. In so doing, we created the opportunity for a methodological experiment as these secondary questionnaires varied in length and relevance, both characteristics known to influence survey response rates [2]. We tested two hypotheses: (1)longer questionnaires (23 or 34 versus 10 items included in secondary outcome measures) will produce lower rates of follow-up and (2)more relevant questionnaires (defined as assessing alcohol problems rather than mental health) will produce higher rates of follow-up.

Methods

Study Procedures and Participants

The methodological studies reported here were embedded in DYD-RCT, a large trial of an online intervention to help hazardous drinkers reduce their alcohol consumption [9]. The parent study included a pilot phase followed by the main trial. The pilot phase involved an unusually large sample, greater than that required for the main trial. We undertook the present methodological experiment in both phases of the parent trial with one alteration made to the design of the second study (see below). We also explored the effects of incentives on follow-up rates in randomized studies, which have been separately reported and do not influence the findings of the present study.

Potential study participants originally accessed a webpage inviting them to "find out if you are drinking too much," and were then asked to complete a brief 3-item screening test. If eligible, they were invited to take part and given access to a consent page after an information page. Eligible participants were people drinking potentially unhealthy levels of alcohol who were also willing to consider changing their behaviour. After a password had been created and email details validated, participants completed the EQ-5D, a well known brief health-related quality of life measure, and calculated their past week alcohol consumption based on specific alcohol brands and volumes. This is a complex task requiring time and effort varying with amount and patterns of drinking. Participants subsequently answered two questions on confidence and intentions before arriving at a *final questionnaire* prior to being told their parent trial group allocation. Without their knowledge, participants had been randomly allocated to one of four different questionnaires (described below) to be completed as this final questionnaire. Participants were thus blinded to the conduct of this study.

All participants thus completed common trial entry and baseline research assessments with the sole difference between the study groups being the secondary outcome measure (ie, the final questionnaire) to which they had been randomly allocated. In both phases, randomization was performed by a computer-generated randomization procedure. Randomization could not be subverted, therefore, by the study team, and allocation was fully concealed. Randomization to a particular secondary outcome measure applied to baseline and both subsequent follow-ups. Participants were thus offered the same secondary outcome questionnaire at all three time points. Randomization was performed separately and independently from randomization to intervention and control conditions in the parent trial [9]. The numbers of participants in the present study slightly exceed those in the parent trial as some participants completed the first randomization to secondary

XSL•FO
RenderX

outcome questionnaire and did not complete the subsequent randomization to parent trial study condition.

Participants were sent email requests for follow-up data in the pilot phase after 1 and 3 months (study 1) and in the main trial phase after 3 and 12 months (study 2). Data collected at follow-up consisted of past week alcohol consumption, the EQ-5D, single-item measures of confidence and intention, and the same secondary outcome measure completed at baseline. Up to three reminders were sent at 7-day intervals to nonresponders, with the final reminder containing a request for participants to tell us their past week alcohol consumption only. Ethical approval was obtained from University College London ethics committee.

Outcomes and Measures

In both studies the sole outcome evaluated here was the proportion of participants who responded, that is, completed the primary outcome questionnaires within 40 days of the email request. The three alcohol problems measures used in both studies were the Alcohol Use Disorders Identification Test (AUDIT), which is the screening test for hazardous and harmful drinking recommended by the World Health Organization [10]; the Leeds Dependence Questionnaire (LDQ), which assesses severity of alcohol dependence [11]; and the Alcohol Problems Questionnaire (APQ), which assesses problems other than dependence [12]. We used the core 23 items of the APQ. The AUDIT and LDQ both comprise 10 items. These instruments all require the respondent to report whether drinking is responsible for a range of difficulties they may experience. Mental health was assessed with different versions of the same instrument: the full 23 or 34 item CORE-OM (Clinical Outcomes in Routine Evaluation-Outcome Measure) in study 1 and the newer, briefer 10-item CORE-10 in study 2 [13-14]. This instrument makes no reference to alcohol.

Statistical Methods

The analyses followed an analysis plan that was written before the relevant data were analyzed. The main analyses compared the proportion responding at each time point between those randomized to longer (APQ and CORE-OM) and shorter (AUDIT, LDQ, and CORE-10) questionnaires and between those randomized to questionnaires relevant to alcohol problems (AUDIT, LDQ, and APQ) and questionnaires less relevant to alcohol problems, being concerned with mental health (CORE-OM and CORE-10). Comparisons were expressed as differences in proportions (risk differences) for interpretability and odds ratios for comparability with other literature.

As a sensitivity analysis, we used logistic regression to adjust for the following baseline variables that were previously found to be predictive of attrition: parent trial group allocation (DYD or comparator), age, gender, educational attainment (degree level or not), ethnicity (white British or other), whether an address was given at study entry, health state, baseline weekly alcohol consumption, and intention (scored 1 to 5). Pooled analyses (that were not prespecified) combined all four follow-up assessments and allowed for the correlation between the two follow-up assessments for the same person using generalized estimating equations adjusting for study and occasion [15]. Prespecified subgroup analyses explored, using interaction tests on both scales, whether any effects differed by gender, parent trial group allocation, educational attainment (university or college degree obtained or not), and baseline weekly alcohol consumption. Baseline weekly alcohol consumption was dichotomized so that women drinking under 35 United Kingdom (UK) units of alcohol and men drinking under 50 UK units of alcohol in the past week were classified as lighter drinkers and those drinking above these levels were deemed heavier drinkers with 1 UK unit being equivalent to 8 grams of ethanol.

Results

Randomization was successful in creating groups equivalent for comparison purposes (Table 1). The total number who consented to participate in the parent trial between February 2007 through August 2008 was 8285 (4957 in study 1 and 3328 in study 2). Of these, 1838 did not complete earlier recruitment steps prior to being randomized to secondary outcome questionnaires, resulting in 6447 study participants for whom results are reported in Table 1. The follow-up rates in groups randomized to the four secondary outcome measures at all four follow-up intervals are presented in Table 2.

XSL•FO
RenderX

Table 1. Baseline characteristics of groups randomized to four secondary outcome measures

	Secondary Outcome Measure					
Baseline Characteristic	AUDIT	LDQ	APQ	CORE-OM[a]	CORE-10[a]	CORE-OM or CORE-10[a]
Number	1614	1607	1613	945	668	1613
Female (%)	58	55	57	55	60	57
Intervention (%)	50	49	50	53	50	52
Heavy drinking (%)	62	62	61	59	62	61
Educated to degree level (%)	50	53	52	50	48	49
White British (%)	84	84	85	84	84	84
Provided postal address (%)	36	34	35	35	35	35
Intentions score, median (interquartile range)	4 (2)	4 (2)	4 (2)	4 (2)	4 (2)	4 (2)
Age, mean (SD)	37.9 (10.7)	37.9 (10.6)	38.0 (10.9)	37.0 (11.0)	37.4 (10.6)	37.2 (10.9)
Health state, mean (SD)	67.4 (22.6)	66.9 (27.1)	67.4 (23.2)	70.8 (46.0)	66.2 (22.9)	68.9 (38.3)
Past week alcohol consumption (UK units), mean (SD)	56.2 (37.9)	57.2 (37.4)	56.8 (40.4)	54.8 (36.3)	56.6 (37.7)	55.5 (36.9)

[a]CORE-OM was used in study 1, CORE-10, in study 2

Table 2. Follow-up rates in groups randomized to four secondary outcome measures

	Secondary Outcome Measure			
Follow-up Period	AUDIT	LDQ	APQ	CORE-OM or CORE-10[a] [1]
Study 1				
1 month	497/949 (52%)	552/939 (59%)	529/947 (56%)	489/945 (52%)
3 months	376/949 (40%)	403/939 (43%)	414/947 (44%)	378/945 (40%)
Study 2				
3 months	337/665 (51%)	316/668 (47%)	316/666 (47%)	308/668 (46%)
12 months	222/665 (33%)	225/668 (34%)	213/666 (32%)	194/668 (29%)

[a]CORE-OM in study 1, CORE-10 in study 2

Shown in Table 3 are comparisons of the follow-up rates between groups randomized to longer (23 or 34 items) and shorter (10 items) secondary outcome measures. Note that the sample sizes are similar in study 1 as there were two questionnaires in each category and dissimilar in study 2 where there was only one longer questionnaire (APQ) and three shorter ones. There is no evidence of any difference in attrition due to additional questionnaire length, and the 95% confidence interval suggests that any difference is no more than 2 percentage points.

XSL•FO
RenderX

Table 3. Follow-up rates in those allocated to longer and shorter secondary outcome questionnaires

	Questionnaire Length				
	Longer	Shorter	Longer vs Shorter		
Follow-up Period			Difference	Odds Ratio	*P Value*
Study 1					
1 month	1018/1892 (54%)	1049/1888 (56%)	−0.02% (−0.05% to 0.01%)	0.93 (0.82-1.06)	.28
3 months	792/1892 (42%)	779/1888 (41%)	0.01% (−0.03% to 0.04%)	1.03 (0.90-1.17)	.71
Study 2					
3 months	316/666 (47%)	961/2001 (48%)	−0.01% (−0.05% to 0.04%)	0.98 (0.82-1.16)	.80
12 months	213/666 (32%)	641/2001 (32%)	−0.00% (−0.04% to 0.04%)	1.00 (0.83-1.20)	.98
Pooled analysis of both studies at all four follow-up intervals			−0.00% (−0.03% to 0.02%)	0.98 (0.89-1.07)	.67
Pooled analysis adjusted for covariates			−0.00% (−0.02% to 0.02%)	0.98 (0.90-1.08)	.75

Data comparing follow-up rates in groups randomized to the three measures of alcohol problems with those randomized to the mental health measure in both study 1 and study 2 are presented in Table 4. The post hoc pooled analysis identifies relevance to alcohol problems to be associated with a 3 percentage point increase in response rate, a result that was clearly statistically significant on unadjusted analysis but only of borderline statistical significance in the sensitivity analysis adjusting for baseline covariates.

Subgroup analyses by the four prespecified covariates identified no strong evidence of effect modification. All *P* values for interaction terms were in excess of .05 whether analyzed separately by study and time (as was prespecified) or pooled over studies and times.

Table 4. Follow-up rates in those allocated to more and less relevant (alcohol problems and mental health respectively) secondary outcome questionnaires

	Questionnaire Focus				
	Alcohol Problems	Mental Health	Alcohol Problems vs Mental Health		
Follow-up Period			Difference	Odds Ratio	*P Value*
Study 1					
1 month	1578/2835 (56%)	489/945 (52%)	0.04% (0.00%-0.08%)	1.17 (1.01-1.36)	.04
3 months	1193/2835 (42%)	378/945 (40%)	0.02% (−0.02% to 0.06%)	1.09 (0.94-1.27)	.26
Study 2					
3 months	969/1999 (48%)	308/668 (46%)	0.02% (−0.02% to 0.07%)	1.10 (0.92-1.31)	.29
12 months	660/1999 (33%)	194/668 (29%)	0.04% (−0.00% to 0.08%)	1.20 (0.99-1.46)	.05
Pooled analysis of both studies at all four follow-up intervals			0.03% (0.01%-0.05%)	1.14 (1.03-1.25)	.01
Pooled analysis adjusted for covariates			0.03% (0.00%-0.05%)	1.11 (1.00-1.22)	.05

Discussion

Allocating participants to longer secondary outcome questionnaires did not lead to lower rates of follow-up when comparing 23 or 34 versus 10 items in addition to completion of primary outcome measures and associated trial entry procedures. More precisely, inspection of the confidence intervals indicates that secondary outcome questionnaire length

XSL•FO
RenderX

does not reduce follow-up rates by more than approximately 2%. More relevant questionnaires assessing alcohol problems rather than mental health did produce higher rates of follow-up though the difference was small, being not greater than 5%, and the statistical significance was doubtful in the sensitivity analysis.These two main findings will first be considered separately.

Questionnaire Length

The unusual decision to randomize to secondary outcome measures was made to minimize attrition, both because we were persuaded by existing high quality evidence of the effects of questionnaire length on response rate and also because attrition was correctly anticipated as a formidable challenge in a trial undertaken completely online. We did not, however, investigate overall assessment burden, which could have been done by making a randomized comparison between the total burden, that is, completion of all secondary outcome measures, which is standard practice, versus one only. This would have required a comparison that assigned a large proportion of participants to a condition expected to be unfavorable to retention in the trial, and, therefore, we chose not to do this. This original study design decision is reemphasised here because of the implications for the interpretation of study findings.

We found that asking participants to answer an additional 23 or 34 questions rather than an additional 10 questions did not influence the likelihood of retention in the study. The unit of analysis in previous postal studies has been the number of pages per questionnaire [2] rather than the number of items per questionnaire, as was used in the present study. In both the previously cited review and in a related systematic review and meta-regression study, Edwards and colleagues identified significant unexplained heterogeneity in the effects of questionnaire length [16]. Effects were greatest when postcards were compared with conventional questionnaires. In six trials comparing one page against either two or three pages, no differences in response rate were observed. Only in the five trials in which a one-page questionnaire was compared against four or more pages did effects on response rate emerge [16].

We are aware of only one previous experimental study in a similar population of drinkers thinking about quitting or reducing their consumption that was not included in previous reviews [17]. It found that a brief alcohol consumption measure yielded a much higher response rate (51%) than did a more detailed and relatively time-consuming measure (22%). Both are commonly used approaches, though the time commitments involved have not been studied.

There are two previous online studies of the effects of questionnaire length on response rate. Both studies found shorter questionnaires to increase response rates by approximately 50% to 100%, which is in line with the mean size of effects observed in postal surveys. Deutskens and colleagues [18] compared a questionnaire taking approximately 15 to 30 minutes to complete with one taking 30 to 45 minutes to complete and found response rates of 24.5% and 17.1% respectively. Marcus and colleagues [19] compared a questionnaire with 91 items taking approximately 10 to 20 minutes to complete with one comprised of 359 items and described as taking 30 to 60 minutes to complete and obtained response rates of 30.8% and 18.6% respectively with the odds of response calculated as 0.51 (95% CI 0.42-0.62). Although the online follow-up context of the present study complicates a direct comparison with the wider literature, it seems very likely that the comparison we made is thus consistent with previous findings in that the difference between the two questionnaire lengths was simply too small to impact upon attrition.

Relevance

Those participants asked more relevant questions in the form of items addressing alcohol problems rather than mental health were on average 3% less likely to be lost to follow-up. These additional questions followed detailed questions about recent alcohol consumption. These findings suggest that the perceived relevance of research assessments could indeed influence attrition.

Our emphasis here is on perceived relevance in the context of an alcohol rather than a mental health trial, even though the perception itself has not been directly assessed. Some participants undoubtedly did have mental health difficulties and may have seen the mental health instrument as being just as relevant to their situation as an alcohol problems measure had they been offered one. Study findings indicate that it is some unspecified property of this instrument that leads to lower follow-up rates in comparison with the others. We assumed at the outset, however, that across the study population as a whole, the mental health content of the additional questionnaire would be viewed as less relevant than an alcohol problems one, and this assumption formed the basis of the hypothesis and the operationalization of the relevance construct. This remains our interpretation of the characteristic most likely to be responsible for the observed difference, though the possibility must be recognized that other features may also be at work.

The existing literature on relevance is rather less extensive than that available for questionnaire length, though again observed effects are much larger than were observed here (unadjusted OR = 1.14, adjusted OR = 1.11). Relevance has also been operationalized heterogeneously in these previous studies. There were three postal studies included in the review by Edwards and colleagues [2] for which the combined odds of response were approximately doubled when more and less relevant questionnaires were compared (OR = 2.00, 95% CI 1.32-3.04). These studies compared the effects on response rates of questionnaires on (1) skipping classes among undergraduates versus PhD students [20], (2) providing bowling versus restaurant feedback among participants in an amateur bowling competition [21], and (3) “a variety of interesting topics” versus “ a boring topic in-depth” in market research [22] with the latter condition in each study being deemed to be less relevant. The only online relevance experiment of which we are aware found a similar effect size to ours (OR = 1.85, 95% CI 1.52-2.26) when comparing a highly *salient* questionnaire on the motives and personality of website owners against one on psychological aspects of Internet usage, which was deemed to be of relatively low salience among website owners [19]. As with the questionnaire length study, although comparisons are necessarily indirect, our relevance experiment involves a much less

XSL•FO
RenderX

pronounced contrast than any previously studied, including this online experiment. We compared two questionnaires both judged relevant to the needs of the study population at the outset though differing in likely degree of relevance as perceived by study participants, whereas in the postal studies, the relevance experiments have been designed to compare relevant with not relevant.

Putting the Findings Together

Our findings are strengthened by the large sample sizes employed, the randomized design, and the absence of any missing data given the nature of the study. The online context of the present study is important, as the Internet is likely to be the vehicle for an increasing number of studies of delivering health care and health promotion in the future, as well as many other types of research. The generalizability of data from this study population of hazardous and harmful drinkers to other populations is unknown.

The original decision to randomize to secondary outcome measures was influenced by the emerging literatures on "assessment reactivity" in the alcohol field [23-24] and on "mere measurement effects" elsewhere [25]. Work in this area suggests that participating in research studies and completing questionnaires can itself influence the target behavior under investigation, which, though relevant to all research designs, could be a particular problem in trials. The present findings on relevance, set in the context of the literature in this area, underscore how little is known about the unintended impacts of our research decisions on participant experience and behavior. The findings also indicate that attrition itself may be a useful proxy measure for unintended adverse effects of research design decisions.

We isolated two aspects of methodological decision making for experimental study here. Qualitative differences in questionnaire content were related to attrition, which suggests the possibilities that the aggregate effects of our methodological decisions may have a large influence not only on attrition but probably also on participant engagement with research in other ways. The absence of an effect of additional questionnaire length on attrition suggests that not all our decisions will do so. This suggestion is coherent with existing online findings of interactions between characteristics affecting response rates in surveys [19].

Further methodological studies of this type are important to pursue specifically in the context of both online and conventional trials and also more broadly, as the lack of prior study of the dynamics of response and attrition in different study designs should be rectified. Surely whether prospective research participants decide to enter studies or not, or stay in them if they do, depends upon what it is that is being asked of them.

Acknowledgments

The authors would like to thank Richard McGregor of Codeface Limited for database programming and IT support and Orla O'Donnell for administrative support. We would also like to thank the National Prevention Research Initiative for funding these studies as part of the pilot and main DYD-RCT. This research was funded by the National Prevention Research Initiative (http://www.npri.org.uk) with support from the following organizations: British Heart Foundation; Cancer Research UK; Chief Scientist Office, Scottish Government Health Directorate; Department of Health; Diabetes UK; Economic and Social Research Council; Health & Social Care Research & Development Office for Northern Ireland; Medical Research Council; Welsh Assembly Government; and World Cancer Research Fund. The funders did not influence any aspect of study design, conduct, analysis, or preparation of manuscript. JM is funded by a Wellcome Trust Research Career Development fellowship in Basic Biomedical Science (WT086516MA). IRW was funded by Medical Research Coucil grant U.1052.00.006.

Conflicts of Interest

None declared

Authors' Contributions

The first author (JM) developed the idea for the study and wrote the first draft of the paper. The second (EK) and third (IW) authors developed the analysis plan in discussion with the first author, and the analysis was done by EK. All authors interpreted the results, contributed to redrafting, and approved the final manuscript. JM is the guarantor.

References

1. Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R, et al. Methods to increase response rates to postal questionnaires. Cochrane Database Syst Rev 2007(2):MR000008. [doi: 10.1002/14651858.MR000008.pub3] [Medline: 17443629]
2. Edwards PJ, Roberts I, Clarke MJ, Diguiseppi C, Wentz R, Kwan I, et al. Methods to increase response to postal and electronic questionnaires. Cochrane Database Syst Rev 2009(3):MR000008. [doi: 10.1002/14651858.MR000008.pub4] [Medline: 19588449]
3. Robinson KA, Dennison CR, Wayman DM, Pronovost PJ, Needham DM. Systematic review identifies number of strategies important for retaining study participants. J Clin Epidemiol 2007 Aug;60(8):757-765. [doi: 10.1016/j.jclinepi.2006.11.023] [Medline: 17606170]

XSL•FO
RenderX

4. Fabricatore AN, Wadden TA, Moore RH, Butryn ML, Gravallese EA, Erondu NE, et al. Attrition from randomized controlled trials of pharmacological weight loss agents: a systematic review and analysis. Obes Rev 2009 May;10(3):333-341. [doi: 10.1111/j.1467-789X.2009.00567.x] [Medline: 19389060]
5. Heo M, Papademetriou E, Meyers BS. Design characteristics that influence attrition in geriatric antidepressant trials: meta-analysis. Int J Geriatr Psychiatry 2009 Sep;24(9):990-1001. [doi: 10.1002/gps.2211] [Medline: 19263464]
6. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res 2009;11(2):e9 [FREE Full text] [doi: 10.2196/jmir.1052] [Medline: 19403465]
7. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
8. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. Lancet 2002 Mar 2;359(9308):781-785. [doi: 10.1016/S0140-6736(02)07882-0] [Medline: 11888606]
9. Murray E, McCambridge J, Khadjesari Z, White IR, Thompson SG, Godfrey C, et al. The DYD-RCT protocol: an on-line randomised controlled trial of an interactive computer-based intervention compared with a standard information website to reduce alcohol consumption among hazardous drinkers. BMC Public Health 2007;7:306 [FREE Full text] [doi: 10.1186/1471-2458-7-306] [Medline: 17963483]
10. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [Medline: 8329970]
11. Raistrick D, Bradshaw J, Tober G, Weiner J, Allison J, Healey C. Development of the Leeds Dependence Questionnaire (LDQ): a questionnaire to measure alcohol and opiate dependence in the context of a treatment evaluation package. Addiction 1994 May;89(5):563-572. [Medline: 8044122]
12. Williams BT, Drummond DC. The Alcohol Problems Questionnaire: reliability and validity. Drug Alcohol Depend 1994 Jun;35(3):239-243. [Medline: 7956754]
13. Barkham M, Margison F, Leach C, Lucock M, Mellor-Clark J, Evans C, et al. Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. Clinical Outcomes in Routine Evaluation-Outcome Measures. J Consult Clin Psychol 2001 Apr;69(2):184-196. [Medline: 11393596]
14. Gilbody S, Richards D, Barkham M. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. Br J Gen Pract 2007 Aug;57(541):650-652. [Medline: 17688760]
15. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986 Mar;42(1):121-130. [Medline: 3719049]
16. Edwards P, Roberts I, Sandercock P, Frost C. Follow-up by mail in clinical trials: does questionnaire length matter? Control Clin Trials 2004 Feb;25(1):31-52. [doi: 10.1016/j.cct.2003.08.013] [Medline: 14980747]
17. Cunningham JA, Ansara D, Wild TC, Toneatto T, Koski-Jännes A. What is the price of perfection? The hidden costs of using detailed assessment instruments to measure alcohol consumption. J Stud Alcohol 1999 Nov;60(6):756-758. [Medline: 10606486]
18. Deutskens E, DeRuyter, K, Wetzels, M, Oosterveld, P. : Response rate and response quality of internet-based surveys: an experimental study. Marketing Letters 2004;15(1):21-36.
19. Marcus B, Bosnjak M, Lindner S, Pilischenko S, Schütz A. Compensating for low topic interest. Social Science Computer Review 2007;25(3):372-338. [doi: 10.1177/0894439307297606]
20. Chen C. Questionnaire length, salience and researchers' authority, and follow-up: the effect on response rates for postal questionnaires. Chinese Journal of Psychology 1984;26(2):77-84.
21. Martin CL. The impact of topic interest on mail survey response behaviour. Journal of the Market Research Society 1994;36(4):327-338.
22. Dommeyer CJ. Does response to an offer of mail survey results interact with questionnaire interest? Journal of the Market Research Society 1985;27(1):27-38.
23. McCambridge J, Day M. Randomized controlled trial of the effects of completing the Alcohol Use Disorders Identification Test questionnaire on self-reported hazardous drinking. Addiction 2008 Feb;103(2):241-248. [doi: 10.1111/j.1360-0443.2007.02080.x] [Medline: 18199302]
24. McCambridge J. [Commentary] Research assessments: instruments of bias and brief interventions of the future? Addiction 2009 Aug;104(8):1311-1312. [doi: 10.1111/j.1360-0443.2009.02684.x] [Medline: 19624324]
25. Godin G, Sheeran P, Conner M, Germain M. Asking questions changes behavior: mere measurement effects on frequency of blood donation. Health Psychol 2008 Mar;27(2):179-184. [doi: 10.1037/0278-6133.27.2.179] [Medline: 18377136]

Abbreviations

APQ: Alcohol Problems Questionnaire
AUDIT: Alcohol Use Disorders Identification Test
CI: confidence interval
CORE-OM: Clinical Outcomes in Routine Evaluation-Outcome Measure

XSL•FO
RenderX

CORE-10: Clinical Outcomes in Routine Evaluation-10-item measure
DYD-RCT: Down Your Drink randomized controlled trial
EQ-5D: trade mark of the EuroQol group see http://www.euroqol.org/
LDQ: Leeds Dependence Questionnaire

Edited by G Eysenbach; submitted 12.01.11; peer-reviewed by P Edwards, R B; comments to author 14.03.11; revised version received 21.03.11; accepted 09.05.11; published 18.11.11.

Please cite as:
McCambridge J, Kalaitzaki E, White IR, Khadjesari Z, Murray E, Linke S, Thompson SG, Godfrey C, Wallace P
Impact of Length or Relevance of Questionnaires on Attrition in Online Trials: Randomized Controlled Trial
J Med Internet Res 2011;13(4):e96
URL: http://www.jmir.org/2011/4/e96/
doi:10.2196/jmir.1733
PMID:22100793

©Jim McCambridge, Eleftheria Kalaitzaki, Ian R. White, Zarnie Khadjesari, Elizabeth Murray, Stuart Linke, Simon G. Thompson, Christine Godfrey, Paul Wallace. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 18.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Methodological Issues in Internet-Mediated Research: A Randomized Comparison of Internet Versus Mailed Questionnaires

Lisa Whitehead[1], BSC(Hons), MA, PhD

Centre for Postgraduate Nursing Studies, University of Otago, Christchurch, New Zealand

Corresponding Author:
Lisa Whitehead, BSC(Hons), MA, PhD
Centre for Postgraduate Nursing Studies
University of Otago
72 Oxford Terrace
Christchurch, 4345
New Zealand
Phone: 64 0064 364 3858
Fax: 64 0064 (0)3 364 0
Email: lisa.whitehead@otago.ac.nz

Abstract

Background: The majority of Internet-mediated studies use measures developed as paper-and-pencil measures or face-to-face-delivered material. Previous research suggests that the equivalence between online and offline measures must be demonstrated rather than assumed.

Objective: The objective of this study was to explore the equivalence 4 measures completed in an online or offline setting.

Methods: A sample of students (n = 1969) was randomly assigned to complete 4 popular scales (the SF-12v2, the Hospital Anxiety and Depression Scale (HADS), the Fatigue Symptom Inventory, and a single-item fatigue measure) either online or by mail survey (pencil and paper). The response rate was 52.51% (n = 1034) and comparable between the online and offline groups.

Results: Significant differences were noted in fatigue levels between the online and offline group (P = .01) as measured by the Fatigue Symptom Inventory, with the online sample demonstrating higher levels of fatigue. Equivalency was noted for the SF-12v2, the Hospital Anxiety and Depression Scale, and the single-item fatigue measure. Internal consistency was high except for the SF-12v2. The SF-12v2 may not be an ideal measure to use for remote administration.

Conclusions: Equivalency of the Hospital Anxiety and Depression Scale (HADS) and the Physical Component Score and Mental Component Score of the SF-12v2 for online and offline data were demonstrated. Equivalency was not demonstrated for the Fatigue Symptom Inventory. Explanations for the difference in fatigue score between the online and offline samples are unclear. Research that seeks to match samples and control for extraneous online and offline variables is called for, along with exploration of factors that may mediate the completion of questionnaires or alter the respondents' relationship with the same, to enhance progress in this area.

(J Med Internet Res 2011;13(4):e109) doi:10.2196/jmir.1593

KEYWORDS

Paper-and-pencil questionnaire; equivalence; questionnaire

Introduction

An Internet-mediated approach to conducting research in the field of health affords researchers a myriad of advantages, including the ability to reach traditionally difficult-to-access groups such as rural populations, people living with illness and disability, and shift workers, and widens geographical access [1,2]. The Internet offers another route of participation in studies for those unable to leave their homes and for those who find reading common forms of print difficult [3]. The Internet may help to defuse embarrassment, feelings of being judged, or shyness [4] and may enhance disclosure [5]. Internet studies present fewer barriers to participation such as keeping appointments or putting a questionnaire in the mail [6].

While there is evidence that online tests can be reliable and valid [7,8], there is also evidence that psychometric properties may change subtly when a test is placed on the Web [9]. The evaluation of a 5-factor personality inventory [10] found that a

XSL•FO
RenderX

small number of the items loaded on different factors (ie, different factors from those they had loaded on in the offline development sample). Inflated results have been noted on the Hospital Anxiety and Depression Scale (HADS) [11] when completed online [12]. Counter to such results, a study of the equivalency of 16 scales noted no significant difference or trends in the completion of the scales [13], and a study of scales used across 16 countries found no discernible differences either [14]. Equivalence of mental health questionnaires (General Health Questionnaire, Symptom Checklist, Medical Outcomes Study Social Support Survey, Perceived Stress Scale, and Utrecht Coping List) found fair to excellent intraclass correlation coefficients (.54–.91) [14].

A key question is, that if indeed differences exist in the distribution of scores generated from online and offline measures, how critical are these? The distribution of scores becomes particularly important if clinical cut-off points are to be generated from the data collection.

The majority of studies that have been conducted in this field have drawn on existing findings, often poorly matched to the online study group and convenience samples. Few studies generate randomized samples assigned to online or offline completion. Issues of sampling bias must be taken into account when interpreting the results of many studies.

The aim of this study was to explore the equivalence of 4 self-report measures administered in an online and offline (paper version) setting.

Methods

Participants

A sample of 2000 students was randomly selected from a database containing all students enrolled at a university (N = 20,688) and then randomly assigned to either the online or offline completion group. This process was undertaken by a biostatistician independent of the study using the randomization feature in Excel 2007 (Microsoft Corporation, Redmond, WA, USA). Of the 2000 students selected, it was established that 31 students had left the university; the final denominator was 1969 (Table 1).

To enhance the response rate, three follow-ups were sent, unless a participant declined to participate (n = 18). The sample closely matched the wider student population by gender, ethnicity, and makeup of home and overseas students (Table 2).

Table 1. Participant response rates to online versus mail questionnaires

	Selected	Left the university	Final denominator	Declined participation	Unable to deliver
Online	1000	14	986	1	23
Mail	1000	17	983	17	32

Table 2. Characteristics of study sample (total sample, online sample, and mail sample) and total study population

Sample	Male		Female		White		Home student		Overseas student	
	n	%	n	%	n	%	n	%	n	%
Total	373	39.4%	594	61.4%	665	68.2%	842	87.3%	124	12.7%
Online	180	38.3%	290	61.7%	319	67.9%	403	85.9%	66	14%
Mail	193	38.8%	304	61.2%	346	69.9%	439	88.6%	58	11%
Study population	8509	42.50%	11,511	57.50%	13,694	68.40%	17,618	88.0%	2402	12.0%

Procedure

The participants who were randomly assigned to participate by mail questionnaire were sent a letter of introduction and the questionnaire to their home address. A stamped, self-addressed envelope was also included, and participants were asked to complete the questionnaire and return this as soon as possible. Two further reminders were sent by mail to those who had not returned a questionnaire 2 weeks after the initial mailing, and then 2 weeks later. No further reminders were issued after this time.

The participants who were randomly assigned to participate by online questionnaire were emailed an invitation to participate. The email contained a Web link that when clicked took the participant to the questionnaire sited on the university's intranet. The questionnaire was not accessible except through the link provided in the email. The questions were presented 6 to a page and in the same order as in the paper questionnaire. Participants were required to complete all questions and to submit each page, which then automatically brought up the next page of questions. Participants were not able to go back and view responses or change these once they had submitted the page.

Data Collection

The questionnaire contained the HADS [11], the SF-12v2 [15,16], a single fatigue item [17], and the Fatigue Symptom Inventory (FSI) [18].

The HADS [11] is a widely used instrument designed to briefly assess anxiety and depression in nonpsychiatric populations. The HADS comprises 14 items, and 2 subscales with 7 items related to anxiety and 7 items to depression.

The SF-12v2 [15,16] is a measure of functional health across 8 domains and is used worldwide. A Physical Component Score

XSL•FO
RenderX

(PCS) and Mental Component Score (MCS) can be calculated from the items.

The single fatigue item from the Zung Self-Rating Depression Scale, "I get tired for no reason," has been used to screen for cancer-related fatigue among 52 patients attending an ambulatory oncology clinic [17]. Sensitivity of 78.95% and specificity of 87.88% were noted when the cut-off point for fatigue was set at 3 ("A good part of the time") and above, when measured against the FSI. Limitations include the generalizability of this scale to detect fatigue in different settings and for different client groups.

The FSI [18] contains 14 items, each with a 10-point scale designed to measure the intensity and frequency of fatigue and its disruptive impact on quality of life. The FSI was developed with a group of patients with breast cancer and a comparison group of healthy people with no history of cancer. The scale has been further used in an outpatient sample of men and women with a variety of cancer diagnoses [18]. The results indicated that the scale was able to discriminate between people with cancer experiencing fatigue and healthy, disease-free controls, supporting the construct validity of the scale. In addition, the instrument was not keyed to a specific illness, although the scale requires further use to validate its applicability to a range of conditions.

Data Analysis

Data were entered into SPSS version 17 (IBM Corporation, Somers, NY, USA). The internal consistency of each measure was explored using the Cronbach alpha coefficient, mean differences were explored using independent *t* tests, and the effect size of any significant differences were explored using the Cohen d.

Ethical Considerations

The proposal was approved by a University of Otago ethics committee. Return of the questionnaire was taken as consent to participate. The data returned were anonymous; the researcher could not trace the student by response.

Results

Response Rate

Of the 2000 students randomly selected for the study, 31 had left the university. The final denominator was 1969 (Table 1). A few students (n = 18) chose not to take part in the study and informed us by return mail. A total of 55 questionnaires were undeliverable. The response rate, based on the final denominator, was 52.51% (n = 1034). The response rate was higher in the online group (n = 536, 54.4%) than in the mail group (n = 498, 50.7%); however, a review of the completion of questions across the questionnaire (Table 3) shows gradual attrition in the online group who completed the questionnaire online. This was not seen in the mail group. The single fatigue item, the last question before the demographic section, had a higher response rate in the mail group (n = 497, 50.6%) than in the email group (n = 472, 47.9%).

Sample Characteristics

The mean age of participants was 24.07 (SD 8.5) years. The mean age was 23.57 (SD 7.63) years among participants who competed the online questionnaire and 24.54 (SD 9.24) years for the mail questionnaire. There were no significant differences between the online and mail questionnaire groups by age, gender, home or overseas student status, or ethnicity.

Internal Consistency of the Measures

The internal consistency of the subscale of each measure was explored for each sample (Table 4). All scales, except for the SF-12v2, demonstrated good internal consistency in both the online and offline setting.

Distribution of Scores on the Single-Item Fatigue Measure

The distribution of scores on the single-item fatigue measure (Table 5) did not differ significantly between the online and mail groups ($c^2_1 = 0.1$, $P = .79$, Cochran-Armitage test for trend).

Mean Difference by Measure for Online and Mail Groups

The mean score on each measure for the online and mail groups was calculated (Table 4) and differences were explored. The only measure on which a statistically significant difference was noted was the FSI interference score. The mean fatigue interference score was higher for the online participants (mean 20.32, SD 14.59) than for the mail group (mean 18.04, SD 14.45; $t_{970} = 2.45$, $P = .01$). The effect size was very small (Cohen d = 0.07) [19].

Table 3. Participant response rates for individual items/scales

	Qu 1[a]		SF-12v2[b]		FSI[c]		HADS[d]		Single item[e]	
	n	%	n	%	n	%	n	%	n	%
Online	536	54.4%	488	49.5%	474	48.1%	472	47.9%	472	47.9%
Mail	498	50.7%	498	50.7%	498	50.7%	498	50.7%	497	50.6%

[a] Start of the questionnaire.

[b] The SF-12v2 measures physical and emotional health.

[c] Fatigue Symptom Inventory.

[d] Hospital Anxiety and Depression Scale.

[e] "I feel tired for no reason."

XSL•FO
RenderX

Table 4. Mean differences between measures

	Online questionnaire			Mail questionnaire			Difference	95% CI[a]	*t*	df	*P* value
	Mean	SD	Cronbach alpha	Mean	SD	Cronbach alpha					
PCS (SF-12v2)[b]	54.13	7.2	.93	54.62	6.7	.66	–0.49	–1.36 to 0.38	–1.11	984	.27
MCS (SF-12v2)[c]	46.04	9.67	.66	46.28	10.02	.67	–0.22	–1.45 to 1.0	–0.36	984	.72
FSI[d] interference score	20.32	14.59	.93	18.04	14.45	.94	2.27	–0.39 to 0.54	2.45	970	.01
Anxiety	6.39	3.68	.80	6.31	3.72	.80	0.07	–0.105 to 0.66	0.31	968	.75
Depression	3.52	3.04	.76	3.24	3.05	.76	0.28	0.45 to 4.09	1.42	968	.16

[a] Confidence interval.

[b] Physical Component Score of the SF-12v2.

[c] Mental Component Score of the SF-12v2.

[d] Fatigue Symptom Inventory.

Table 5. Single-item fatigue measure ("I feel tired for no reason") score

	0 (none or a little of the time)		1 (some of the time)		2 (a good part of the time)		3 (most of the time)		Total
	n	%	n	%	n	%	n	%	
Online	202	42.8%	212	44.9%	38	8%	20	4%	472
Mail	214	43.1%	212	42.7%	52	10%	19	4%	497
Total	416		424		90		39		969

Discussion

Equivalency of the HADS and of the PCS and MCS of the SF-12v2 for online and offline data were demonstrated. The alpha scores for the SF-12v2 PCS scale in the mail group and the MCS scale in both groups were below the normal threshold of acceptability (.7) and indicate some uncertainty around the results of the online–offline comparisons. The SF-12v2 may not be an ideal measure to use for remote administration. The findings mainly supported those of earlier studies that have found no differences between the online and offline setting. Of note, no differences were found for the HADS, where inequivalence had been noted previously [12]. Possible reasons for the equivalence noted in this study (not noted in the previous study) were that participants were recruited from the same source and were randomly allocated to the online or offline group. Equivalency was not demonstrated for the FSI; however, the effect size of the difference in the mean scores on the FSI between the online and offline groups was very small. Explanations for the difference in fatigue score between the online and offline samples are myriad, although no one answer is likely to explain the situation. Computer aversion, computer anxiety, and computer self-efficacy have been proffered as influencing the completion of online questionnaires [9]. It is unlikely that any of these variables affected the completion of the fatigue questionnaire, where differences in the completion of the other measures were not affected, and where computer anxiety is known to be low and computer self-efficacy medium to high among university students [20]. Unlike previous studies reporting differences between data collected online and offline [21,22], the current study employed random sampling, and no obvious differences were observed between the two samples. The question of whether participants were influenced by social desirability in their response remains open; the online results may reflect greater openness to express symptoms, a phenomenon reported by other researchers [5,23], and chronic fatigue has been reported as viewed pejoratively by others [24]. However, given that self-reports of anxiety and depression, both known to be widely stigmatized, were invariant between the two data approaches, this explanation does not hold much weight either.

Questions remain around the ability to transfer an established measure for completion within an online environment without affecting the construct validity of the measure and the distribution of responses. The evidence to support differences between measures completed online and offline is not clear. There is evidence to suggest that the distribution of responses obtained from an online study may not be directly comparable with established norms. Research that seeks to match sample and control populations for extraneous online and offline variables is called for, along with exploration of factors that may mediate the completion of questionnaires or alter the respondents' relationship with the same, if progress in this area is to be made.

XSL•FO
RenderX

Conflicts of Interest

None declared

References

1. Whitehead LC. Methodological and ethical issues in Internet-mediated research in the field of health: an integrated review of the literature. Soc Sci Med 2007 Aug;65(4):782-791. [doi: 10.1016/j.socscimed.2007.03.005] [Medline: 17512105]
2. Mann C, Stewart F. Internet Communication and Qualitative Research: A Handbook for Researching Online. London: Sage Publications; 2000.
3. Pilling D, Barrett P, Floyd M. Disabled People and the Internet: Experiences, Barriers and Opportunities. London: Joseph Rowntree Foundation; 2004.
4. Valaitis RK, Sword WA. Online discussions with pregnant and parenting adolescents: perspectives and possibilities. Health Promot Pract 2005 Oct;6(4):464-471. [doi: 10.1177/1524839904263897] [Medline: 16210689]
5. Joinson A. Social desirability, anonymity, and Internet-based questionnaires. Behav Res Methods Instrum Comput 1999 Aug;31(3):433-438. [Medline: 10502866]
6. Shih TH, Fan X. Comparing response rates from Web and mail surveys: a meta-analysis. Field Methods 2008;20(3):249-271. [doi: 10.1177/1525822X08317085]
7. Davis RN. Web-based administration of a personality questionnaire: comparison with traditional methods. Behav Res Methods Instrum Comput 1999 Nov;31(4):572-577. [Medline: 10633976]
8. Buchanan T. Potential of the Internet for personality research. In: Birnbaum MD, editor. Psychological Experiments on the Internet. San Diego, CA: Academic Press; 2000.
9. Schulenberg SE, Yutrzenka BA. The equivalence of computerized and paper-and-pencil psychological instruments: implications for measures of negative affect. Behav Res Methods Instrum Comput 1999 May;31(2):315-321. [Medline: 10495816]
10. Buchanan T, Johnson, Goldberg L. Implementing a five-factor personality inventory for use on the Internet. Eur J Psychol Assess 2005;21(2):115-127. [doi: 10.1027/1015-5759.21.2.115]
11. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
12. McCue P, Buchanan T, Martin CR. Screening for psychological distress using internet administration of the Hospital Anxiety and Depression Scale (HADS) in individuals with chronic fatigue syndrome. Br J Clin Psychol 2006 Nov;45(Pt 4):483-498. [doi: 10.1348/014466505X82379] [Medline: 17076959]
13. 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] [doi: 10.2196/jmir.6.3.e29] [Medline: 15471755]
14. De Beuckelaer A, Lievens F. Measurement equivalence of paper-and-pencil and Internet organisational surveys: a large scale examination in 16 countries. Appl Psychol 2009;58(2):336-361. [doi: 10.1111/j.1464-0597.2008.00350.x]
15. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [Medline: 8628042]
16. Ware JE, Kosinski M, Dewey JE. How to Score Version 2 of the SF-12 Health Survey. Boston, MA: Health Assessment Lab; 2002.
17. Kirsh KL, Passik S, Holtsclaw E, Donaghy K, Theobald D. I get tired for no reason: a single item screening for cancer-related fatigue. J Pain Symptom Manage 2001 Nov;22(5):931-937. [Medline: 11728796]
18. Hann DM, Jacobsen PB, Azzarello LM, Martin SC, Curran SL, Fields KK, et al. Measurement of fatigue in cancer patients: development and validation of the Fatigue Symptom Inventory. Qual Life Res 1998 May;7(4):301-310. [Medline: 9610214]
19. Ellis PD. Hong Kong Polytechnic University. 2009. Effect Size Calculators URL: http://www.polyu.edu.hk/mm/effectsizefaqs/calculator/calculator.html [accessed 2011-10-18] [WebCite Cache ID 62WpX3F3y]
20. Sam HK, Othman AEA, Nordin ZS. Computer self-efficacy, computer anxiety, and attitudes toward the Internet: a study among undergraduates in Unimas. J Educ Technol Soc 2005;8(4):205-219.
21. Davidov E, Depner F. Testing for measurement equivalence of human values across online and paper-and-pencil surveys. Qual Quant 2009;45(2):375-390. [doi: 10.1007/s11135-009-9297-9]
22. Ployhart R, Weekley J, Holtz B, Kemp C. Web-based and paper-and-pencil testing of applicants in a proctored setting: are personality, biodata, and situational judgment tests comparable? Pers Psychol 2003;56(3):733-752. [doi: 10.1111/j.1744-6570.2003.tb00757.x]
23. Reitz R, Wahl S. Comparison of psychologist's self-esteem and their image in the internet and in print. In: Batinic B, Reips UD, Bosnjak M, editors. Online Social Sciences. Seattle, WA: Hogrefe & Huber; 2002.
24. Schlossberg D. In: Schlossberg D, editor. Clinical Infectious Disease. Cambridge: Cambridge University Press; 2008.

XSL•FO
RenderX

Abbreviations

FSI: Fatigue Symptom Inventory
HADS: Hospital Anxiety and Depression Scale
MCS: Mental Component Score
PCS: Physical Component Score

Edited by G Eysenbach; submitted 25.06.10; peer-reviewed by T Buchanan, F Holländare; comments to author 06.11.10; revised version received 22.08.11; accepted 25.08.11; published 12.12.11.

Please cite as:
Whitehead L
Methodological Issues in Internet-Mediated Research: A Randomized Comparison of Internet Versus Mailed Questionnaires
J Med Internet Res 2011;13(4):e109
URL: http://www.jmir.org/2011/4/e109/
doi:10.2196/jmir.1593
PMID:22155721

©Lisa Whitehead. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 12.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Internet Use for Health-Related Information via Personal Computers and Cell Phones in Japan: A Cross-Sectional Population-Based Survey

Yoshimitsu Takahashi[1], PhD; Tomoko Ohura[1,2], OTR, MPH; Tatsuro Ishizaki[3], MPH, MD, PhD; Shigeru Okamoto[1], MD; Kenji Miki[4], MPH; Mariko Naito[5], DDS, PhD; Rie Akamatsu[6], DrPH; Hiroki Sugimori[7], MMedSc, MD, PhD; Nobuo Yoshiike[8], MD; Koichi Miyaki[9], MD, PhD; Takuro Shimbo[9], MD; Takeo Nakayama[1], MD, PhD

[1]Department of Health Informatics, Kyoto University School of Public Health, Kyoto, Japan

[2]Faculty of Care and Rehabilitation, Seijoh University, Tokai, Aichi, Japan

[3]Human Care Research Team, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan

[4]Regional Research Institute, Osaka University of Economics and Law, Yao, Osaka, Japan

[5]Department of Preventive Medicine, Graduate School of Medicine, Nagoya University, Nagoya, Aichi, Japan

[6]Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo, Japan

[7]Department of Preventive Medicine, Graduate School of Sports and Health Sciences, Daito Bunka University, Higashimatsuyama, Saitama, Japan

[8]Community Nutrition, Faculty of Health Sciences, Aomori University of Health and Welfare, Aomori, Japan

[9]Department of Clinical Research and Informatics, National Center for Global Health and Medicine, Tokyo, Japan

Corresponding Author:
Takeo Nakayama, MD, PhD
Department of Health Informatics
Kyoto University School of Public Health
Yoshida Konoe
Sakyo
Kyoto, 606-8501
Japan
Phone: 81 75 753 4488
Fax: 81 75 753 4497
Email: nakayama.t@at2.ecs.kyoto-u.ac.jp

Abstract

Background: The Internet is known to be used for health purposes by the general public all over the world. However, little is known about the use of, attitudes toward, and activities regarding eHealth among the Japanese population.

Objectives: This study aimed to measure the prevalence of Internet use for health-related information compared with other sources, and to examine the effects on user knowledge, attitudes, and activities with regard to Internet use for health-related information in Japan. We examined the extent of use via personal computers and cell phones.

Methods: We conducted a cross-sectional survey of a quasi-representative sample (N = 1200) of the Japanese general population aged 15–79 years in September 2007. The main outcome measures were (1) self-reported rates of Internet use in the past year to acquire health-related information and to contact health professionals, family, friends, and peers specifically for health-related purposes, and (2) perceived effects of Internet use on health care.

Results: The prevalence of Internet use via personal computer for acquiring health-related information was 23.8% (286/1200) among those surveyed, whereas the prevalence via cell phone was 6% (77). Internet use via both personal computer and cell phone for communicating with health professionals, family, friends, or peers was not common. The Internet was used via personal computer for acquiring health-related information primarily by younger people, people with higher education levels, and people with higher household incomes. The majority of those who used the Internet for health care purposes responded that the Internet improved their knowledge or affected their lifestyle attitude, and that they felt confident in the health-related information they obtained from the Internet. However, less than one-quarter thought it improved their ability to manage their health or affected their health-related activities.

XSL•FO
RenderX

Conclusions: Japanese moderately used the Internet via personal computers for health purposes, and rarely used the Internet via cell phones. Older people, people with lower education levels, and people with lower household incomes were less likely to access the Internet via cell phone. The Internet moderately improved users' health-related knowledge and attitudes but seldom changed their health-related abilities and activities. To encourage communication between health providers and consumers, it is important to improve eHealth literacy, especially in middle-aged people. It is also important to make adequate amendments to the reimbursement payment system and nationwide eHealth privacy and security framework, and to develop a collaborative relationship among industry, government, and academia.

(J Med Internet Res 2011;13(4):e110) doi:10.2196/jmir.1796

KEYWORDS

eHealth; email; cell phones; health literacy; information-seeking behavior; patient-provider communication

Introduction

The number of Internet users has increased considerably worldwide [1]. The Internet is used for health purposes by the general public, and the importance of the Internet as a source of health information is growing [2-5]. The term eHealth refers to health services and information delivered or enhanced through the Internet [1,6-8]. To monitor health consumers' use, attitudes, and activities regarding health-related information and eHealth, national representative surveys were conducted in the United States (Health Information National Trends Survey by the National Cancer Institute) [3,9-11] and Europe (eHealth Consumer Trends Survey funded by the European Commission) [5,12,13]. These surveys revealed an increase in the prevalence of Internet use for health-related information among the general public. The prevalence in the United States was approximately 20% in 2001 [2] and 40% in 2003 [3]; the prevalence in Europe was 42% in 2005 and 52% in 2007 [5]. Several studies, however, showed that people still valued and used more conventional sources of health-related information, including health professionals, family, television, and newspapers, although the conventional sources decreased in importance [14,15]. It was also shown that the effects of Internet use on health-related attitudes and activities, such as active communication and actual health care utilization, have not yet been substantial [2,5].

A 2007 Japanese national survey showed that 69% used the Internet in the past year, 61% through personal computers and 57% through cell phones [16]. Japan's cell phones are technologically enhanced and divergent from globalization, a phenomenon labeled the Galápagos syndrome [17]. They are ready for the Internet and email, have high-resolution cameras, receive television programs, and can be used as credit cards and boarding passes. Even the average person can have an advanced cell phone, so many Japanese rely on their cell phones rather than personal computers for Internet access [17]. However, little is known about the use of, attitudes toward, and activities regarding eHealth in the Japanese population. Clearer fundamental information is required as a foundation for discussing the role of the Internet in health care. It is assumed that changes in the information technology environment have affected Internet use for health-related information in several ways in Japan, as well as in the United States and Europe, where Web 2.0 has been changing the way medical information is handled (eg, personal health records) [18,19]. Therefore, we hypothesized that the prevalence and effects of Internet use for health-related information in Japan are similar to those in the United States and Europe. Moreover, since many Japanese rely on their cell phones for Internet access, we think that Internet use via cell phone can be as effective as Internet use via personal computer [16,17,20].

This study aimed to measure the prevalence of Internet use for health-related information compared with other sources, to examine user characteristics, and to examine the association of Internet use with user knowledge, attitudes, and activities regarding health-related information in Japan. Additionally, we examined the extent of Internet use via personal computers and cell phones.

Methods

Study Design and Participants

We designed a cross-sectional survey of the Japanese general population aged 15–79 years. We used a scheduled omnibus survey conducted by Nippon Research Center Ltd [21], which included 1200 participants. Study participants were selected by proportional quota sampling to collect a nationally representative sample in Japan, and a self-reported questionnaire survey was performed in September 2007. In proportion to regions and city sizes, 200 areas were proportionately selected corresponding to the stratification of all nine regions (Hokkaido, Tohoku, Kanto, Hokuriku, Tokai, Kinki, Chugoku, Shikoku, and Kyushu) and five city sizes (15 large cities: Sapporo, Sendai, Saitama, Chiba, Tokyo, Kawasaki, Yokohama, Shizuoka, Nagoya, Kyoto, Osaka, Kobe, Hiroshima, Kitakyushu, and Fukuoka; cities with over 150,000 people; cities with over 50,000 people; cities with fewer than 50,000 people; and nonurban areas). Households were randomly selected from a database of house maps. Individuals were allocated to reflect the area's stratification by sex, age, and job status. Interviewers visited selected households, requested that individuals fill out questionnaires, and collected questionnaires completed by allocated individuals a few days later. In the case that interviewers could not collect a questionnaire from a target participant, interviewers visited the next target participant that reflected the area's demographics. Sampling continued until we had 1200 completed questionnaires. All respondents were provided ¥1000 (about US $10 at the time of writing) as payment on completion of the questionnaire.

XSL•FO
RenderX

Measurements

The survey contained a set of questions about participant characteristics, use of the Internet for health-related information, and the perceived effects of Internet use on knowledge, attitudes, and activities for health purposes. Almost all items were derived from the original questionnaire used in Baker and colleague's study [2]. We added some original items regarding cell phones.

We collected basic demographic data from participants, including age, sex, household income, level of education, and place of residence. Health-related characteristics were self-reported health status (excellent, very good, good, fair, or poor) and chronic diseases: hypertension, diabetes or hyperglycemia, cancer, heart problems (heart attack, angina due to coronary heart disease, heart failure, or other heart problems), depression, obesity, and hyperlipidemia. The main outcome of this study was frequency of Internet use for any purpose and ownership of cell phones.

We classified Internet use into four types: (1) use of a Web browser via personal computer, (2) use of a Web browser via cell phone, (3) use of email via personal computer, and (4) use of email via cell phone. We prepared four questions: "How often do you use a Web browser (or email) to acquire information or advice for health care via your personal computer (or through your cell phone)?" We defined "Internet use" as more than once a year. In addition, to compare the extent of Internet use, we also measured the extent other sources were used for health-related information (television, newspapers, radio, magazines, direct mail, and public relations magazines). To investigate the extent of interactive Internet use for health-related communication, we asked participants about their use of the Internet for three purposes: "to contact doctors or other health care providers," "to contact a family member or friend about health or health care," and "to contact other people who have similar health conditions or concerns." We examined the extent of use for these three purposes via personal computer and cell phone.

We examined the perceived effects of Internet use on knowledge and attitudes using participant responses (strongly agree, agree, disagree, or strongly disagree) to the following statements: "improved my understanding of symptoms, conditions, or treatments in which I was interested," "improved my ability to manage my health care needs without visiting a doctor or other health care provider," "led me to seek care from different doctors or health providers than I otherwise would have," and "affected the way I eat or exercise." We also examined Internet user confidence or anxiety ("I felt confident," "I wasn't influenced," "I felt anxious," or "I've never obtained [this information]" after obtaining the following health-related information: "information on diseases you have," "information on diseases you want to prevent," "information on treatment of diseases," "information on doctors and health care facilities," "information on peers," and "information on a healthy lifestyle, fitness, or nutrition."

We examined the perceived effects of Internet use on activities by collecting data on the number of times participants visited a health professional and the number of times they telephoned them. Additionally, we asked "Have you ever told health professionals about information from the Internet?"

Statistical Analysis

We tabulated the responses and computed the prevalences. Then, we used logistic regression analysis to investigate the relationships between Internet use for health-related information and respondent characteristics (age, sex, annual household income, level of education, place of residence, and self-reported health status). We evaluated eight logistic regression models. The outcomes for models 1–4 were use of the Internet via personal computer for (1) acquiring health-related information, (2) contacting health professionals, (3) contacting family/friends about health-related information, and (4) contacting peers. The outcomes for models 5–8 were use of the Internet via cell phone for (5) acquiring health-related information, (6) contacting health professionals, (7) contacting family/friends, and (8) contacting peers. For each variable, we report odds ratios and 95% confidence intervals. The Hosmer-Lemeshow goodness-of-fit test was performed for each model. All analyses were performed using SPSS version 18.0 (IBM Corporation, Somers, NY, USA). All P values were 2-sided, with $P < .05$ considered statistically significant.

Ethical Considerations

The purpose of the study was explained on the first page of the questionnaire, and we declared that responses to questionnaires were regarded as informed consent. This survey was conducted as an unlinked anonymous survey. The study protocol was approved by the Ethics Committee of Kyoto University Faculty of Medicine.

Results

Participant Characteristics

Characteristics of the 1200 survey participants included in our analysis are shown in Table 1. The mean (SD) age was 46.4 (17.4) years, 49.6% (595/1200) of the participants were male, 18.7% (224) had at least a college education, and 35.6% (426) had a household income of ¥6,000,000 (about US $60,000) or more. "Poor" general health was reported by 7% (82) of respondents and 31.4% (377) had at least one chronic condition. In addition, 41.5% (498) had used the Internet more than once a week for general purposes; 81.3% (975) had cell phones, and the prevalence of Internet use via cell phone was 41.8% (502).

XSL•FO
RenderX

Table 1. Participant characteristics (N = 1200)

	n	%
Age (years)		
15–19	75	6
20–34	285	23.8
35–49	295	24.6
50–64	324	27.0
65–74	169	14.1
75–79	52	4
Mean (SD)	46.4 (17.4)	
Sex (male)	595	49.6
Household income (¥1000) [a]		
0–2999	194	16.2
3000–5999	418	34.8
6000–9999	314	26.2
10,000–	112	9.3
Unknown	162	13.5
Education (years)		
0–12	728	60.7
13–15	241	20.1
16–	224	18.7
Unknown	7	1
Health status		
Excellent/very good	248	20.7
Good	350	29.2
Fair	520	43.3
Poor	82	7
Place of residence		
Urban[b]	690	57.5
Nonurban	510	42.5
Chronic conditions		
≥3	33	3
2	86	7
1	258	21.5
0	823	68.6

[a] ¥1000 = about US $10.

[b] Cities with a population of at least 150,000 people.

Prevalence of Using Various Sources for Acquiring Health-Related Information

Table 2 shows the prevalence of use for health-related information by source. We regarded *use* as use at least once every year. The prevalence of Internet use (Web browser or email) via personal computers for acquiring health-related information was 23.8% (286), and 6% (77) for Internet use via cell phones. Television (60.1%, 721) and newspapers (50.3%, 604) were widely used. The prevalence was 7% (79) and 3% (36) for contact with health professionals (doctors or other health care providers), 8.6% (103) and 12.3% (148) for contact with family or friends, and 4% (52) and 6% (67) for contact with peers (other people with similar health conditions or concerns) for interactive use of the Internet via personal computer and cell phone, respectively.

XSL•FO
RenderX

Table 2. Prevalence and frequency of Internet use for health purposes (N = 1200)

In the past year, about how often did you	Frequency of use, % (n)					
	Total ever in the past year	More than once per week	About once per week	About once per month	Every 2–3 months	Less than every 2–3 months
To acquire health-related information, use						
Television	60.1 (721)	21.7 (260)	14.8 (178)	10.8 (129)	5 (54)	8.3 (100)
Newspapers	50.3 (604)	17.3 (207)	13.3 (159)	9.1 (109)	4 (48)	7 (81)
Public relations magazines	40.3 (484)	1 (17)	2 (27)	20.6 (247)	6 (77)	9.7 (116)
Magazines	34.2 (410)	3 (38)	4 (50)	12.3 (147)	7 (83)	8 (92)
Radio	19.1 (229)	5 (56)	4 (52)	4 (50)	2 (26)	4 (45)
Direct mail	16.5 (198)	1 (15)	2 (25)	6 (68)	4 (45)	4 (45)
Web browser via...						
Personal computer	23.7 (284)	4 (45)	4 (47)	6 (69)	6 (67)	5 (56)
Cell phone	5 (63)	1 (14)	1 (8)	1 (15)	1 (8)	2 (18)
Email via...						
Personal computer	5 (61)	1 (15)	1 (8)	1 (15)	1 (8)	1 (15)
Cell phone	4 (48)	1 (14)	0 (6)	1 (12)	0 (2)	1 (14)
The Internet (Web browser or email) via...						
Personal computer	23.8 (286)	4 (48)	4 (47)	6 (70)	5 (65)	5 (56)
Cell phone	6 (77)	2 (21)	1 (7)	2 (19)	1 (7)	2 (23)
To contact health professionals, use						
The Internet via...						
Personal computer	7 (79)	1 (8)	1 (6)	2 (18)	1 (17)	3 (30)
Cell phone	3 (36)	2 (7)	0 (2)	1 (6)	1 (6)	1 (15)
To contact a family member or friend about health or health care, use						
The Internet via...						
Personal computer	8.6 (103)	2 (20)	1 (16)	2 (20)	2 (19)	2 (28)
Cell phone	12.3 (148)	2 (29)	2 (25)	3 (34)	1 (15)	4 (45)
To contact peers about health or health care, use						
The Internet via...						
Personal computer	4 (52)	1 (14)	0 (4)	1 (8)	1 (8)	2 (19)
Cell phone	6 (67)	1 (8)	1 (10)	1 (17)	1 (10)	2 (22)

Characteristics of People Using the Internet for Health-Related Information

Table 3 shows logistic regression analysis results of the relationships between Internet use via personal computer for each health purpose and participant characteristic. We observed that participants over 50 years of age were significantly less likely to use the Internet via personal computer for acquiring health-related information, while those with an income over ¥10,000,000 or with more than 12 years of education were more likely to acquire information this way. Table 4 shows results on the use of cell phones. Participants aged 35–64 years were less likely than younger participants to use the Internet via cell phone to obtain health-related information.

We also considered interactive use of the Internet for health-related communication (see Table 3 and Table 4). Overall, there were few statistically significant differences in characteristics. Women and those reporting good health status were more likely to use the Internet interactively. As with using the Internet via personal computer for acquiring information, higher rates of interactive Internet use might be related to younger age, higher education levels, and higher household incomes. A lower rate of interactive Internet use via cell phone might be associated with older age and lower education levels, but was not associated with household income. All eight models were shown to be well calibrated using the Hosmer-Lemeshow test (each P value > .31).

XSL•FO
RenderX

Table 3. Results of logistic regression models for Internet use via personal computer for each health purpose by demographic characteristics (N = 1200)

	For acquiring information	For contacting professionals	For contacting family/friends	For contacting peers
Number of users (%)	286 (23.8%)	79 (7%)	103 (8.6%)	52 (4%)
Age (years), OR (95% CI)[a]				
15–19	0.5 (0.2–1.0)	0.2 (0.0–1.3)	0.6 (0.2–1.7)	0.3 (0.0–2.5)
20–34	Reference	Reference	Reference	Reference
35–49	1.2 (0.8–1.7)	1.0 (0.6–1.7)	0.9 (0.5–1.4)	0.8 (0.4–1.6)
50–64	0.6 (0.4–0.9)[b]	0.7 (0.4–1.4)	0.7 (0.4–1.2)	0.7 (0.3–1.5)
65–79	0.2 (0.1–0.4)[b]	0.4 (0.2–1.1)	0.3 (0.1–0.8)[b]	0.3 (0.1–1.1)
Sex (female), OR (95% CI)[a]	1.0 (0.8–1.4)	1.8 (1.1–2.9)[b]	1.5 (0.9–2.3)	1.7 (0.9–3.2)
Household income (¥1000)[c], OR (95% CI)[a]				
0–2999	Reference	Reference	Reference	Reference
3000–5999	1.6 (0.9–2.7)	1.2 (0.5–3.0)	1.1 (0.5–2.5)	1.2 (0.4–3.4)
6000–9999	1.7 (1.0–2.9)	1.6 (0.7–4.0)	1.5 (0.7–3.3)	1.2 (0.4–3.6)
10,000–	2.5 (1.3–4.8)[b]	1.1 (0.4–3.4)	1.6 (0.7–4.0)	2.3 (0.7–7.3)
Education (years), OR (95% CI)[a]				
0–12	Reference	Reference	Reference	Reference
13–15	1.8 (1.2–2.6)[b]	1.9 (1.0–3.3)[b]	1.7 (0.9–2.9)	1.7 (0.8–3.6)
16–	4.8 (3.3–6.8)[b]	2.6 (1.4–4.7)[b]	3.8 (2.3–6.4)[b]	2.8 (1.4–5.8)[b]
Urban residence[d], OR (95% CI)[a]	1.4 (0.9–2.1)	0.6 (0.2–1.4)	1.1 (0.6–2.0)	0.7 (0.3–1.8)
Health status, OR (95% CI)[a]				
Excellent/very good	Reference	Reference	Reference	Reference
Good	1.4 (0.9–2.1)	1.0 (0.5–1.8)	1.3 (0.8–2.3)	3.0 (1.2–7.7)[b]
Fair	1.1 (0.7–1.6)	0.9 (0.5–1.6)	0.7 (0.4–1.2)	1.9 (0.7–4.9)
Poor	1.8 (0.9–3.6)	0.6 (0.2–2.3)	1.1 (0.4–3.1)	1.4 (0.3–7.6)
Test for goodness-of-fit[e]	P = .39	P = .66	P = .99	P = .59

[a] Odds ratio (95% confidence interval).

[b] Confidence interval does not include 1.0.

[c] ¥1000 = about US $10.

[d] Population of at least 150,000 people.

[e] Hosmer-Lemeshow goodness-of-fit test.

XSL•FO
RenderX

Table 4. Results of logistic regression models for Internet use via cell phone for each health purpose by demographic characteristics (N = 1200)

	For acquiring information	For contacting professionals	For contacting family/friends	For contacting peers
Number of users (%)	63 (5%)	36 (3%)	148 (12.3%)	67 (6%)
Age (years), OR (95% CI)[a]				
15–19	0.6 (0.2–1.6)	0.9 (0.3–3.1)	0.6 (0.3–1.2)	0.6 (0.2–1.9)
20–34	Reference	Reference	Reference	Reference
35–49	0.5 (0.3–1.0)[b]	0.3 (0.1–0.8)[b]	0.5 (0.3–0.8)[b]	0.6 (0.4–1.2)
50–64	0.2 (0.1–0.4)[b]	0.2 (0.1–0.6)[b]	0.2 (0.1–0.3)[b]	0.2 (0.1–0.5)[b]
65–79	n/a[c]	n/a[c]	0.0 (0.0–0.1)[b]	0.0 (0.0–0.3)[b]
Sex (female), OR (95% CI)[a]	1.2 (0.7–2.0)	1.3 (0.6–2.7)	2.0 (1.4–3.0)[b]	2.2 (1.3–3.9)[b]
Household income (¥1000)[d], OR (95% CI)[a]				
0–2999	Reference	Reference	Reference	Reference
3000–5999	1.2 (0.5–3.1)	0.6 (0.2–1.9)	0.7 (0.4–1.4)	1.0 (0.4–2.4)
6000–9999	1.2 (0.4–3.2)	1.2 (0.4–3.5)	1.1 (0.6–2.0)	0.9 (0.4–2.4)
10,000–	1.7 (0.6–5.4)	0.7 (0.2–3.3)	1.0 (0.5–2.2)	1.0 (0.3–3.2)
Education (years), OR (95% CI)[a]				
0–12	Reference	Reference	Reference	Reference
13–15	1.1 (0.6–2.2)	1.3 (0.5–3.0)	1.2 (0.8–2.0)	1.2 (0.6–2.3)
16–	1.4 (0.7–2.8)	1.3 (0.5–3.1)	2.0 (1.3–3.2)[b]	1.9 (1.0–3.6)
Urban residence[e], OR (95% CI)[a]	1.9 (1.0–3.8)	0.9 (0.3–2.7)	0.8 (0.4–1.5)	0.5 (0.2–1.4)
Health status, OR (95% CI)[a]				
Excellent/very good	Reference	Reference	Reference	Reference
Good	1.0 (0.5–2.1)	2.2 (0.8–6.0)	1.9 (1.2–3.2)[b]	3.8 (1.7–8.7)[b]
Fair	0.9 (0.4–1.8)	1.5 (0.5–4.2)	1.3 (0.8–2.2)	2.0 (0.8–4.6)
Poor	3.3 (1.1–9.6)	2.5 (0.5–13.5)	1.8 (0.7–4.5)	1.6 (0.3–8.2)
Test for goodness-of-fit[f]	*P* = .99	*P* = .31	*P* = .59	*P* = .31

[a] Odds ratio (95% confidence interval).

[b] Confidence interval does not include 1.0.

[c] Not applicable.

[d] ¥1000 = about US $10.

[e] Population of at least 150,000 people.

[f] Hosmer-Lemeshow goodness-of-fit test.

Perceived Effects of Internet Use on Health Care

Table 5, Table 6, Table 7, and Table 8 show results of perceived effects of Internet use on health care. More than two-thirds of Internet users strongly agreed or agreed that Internet use "improved my understanding of symptoms, conditions, or treatments in which I was interested" (143/210, 68.1%) and "affected the way I eat or exercise" (134/197, 68.0%), while only 23% thought it "improved my ability to manage my health care needs without visiting a doctor or other health care provider." More than 60% of respondents obtaining any kind of health-related information felt confident after obtaining this information. Most respondents thought that Internet use had no effect on the number of times they visited health professionals (208/234, 88.9%) or telephoned health professionals (216/232, 93.1%), and most had never told health professionals about information they obtained from the Internet (197/236, 83.5%).

XSL•FO
RenderX

Table 5. Perceived effects of Internet use on health care understanding and decisions among Internet users

	n	Agree or strongly agree
Improved my understanding of symptoms, conditions, or treatments in which I was interested	210	143 (68.1%)
Affected the way I eat or exercise	197	134 (68.0%)
Led me to seek care from different doctors or health providers than I otherwise would have	190	41 (22%)
Improved my ability to manage my health care needs without visiting a doctor or other health care provider	188	43 (23%)

Table 6. Perceived effects of Internet use on feelings of confidence and anxiety among Internet users

Feeling after obtaining information on...	n	Feeling confident	No effect	Feeling anxious
Diseases you have	158	98 (62%)	52 (33%)	8 (5%)
Diseases you want to prevent	125	77 (62%)	46 (37%)	2 (2%)
Treatment of diseases	167	108 (64.7%)	53 (32%)	6 (4%)
On doctors and health care facilities	99	93 (63%)	50 (34%)	4 (3%)
On peers	147	61 (62%)	34 (34%)	4 (4%)
On a healthy lifestyle, fitness, or nutrition	129	82 (64%)	45 (35%)	2 (2%)

Table 7. Perceived effects of Internet use on health-related activities (number of times visited or telephoned a physician or other health provider) among Internet users

Number of times...	n	Increased	No effect	Decreased
Visited a physician or other health provider	234	15 (6%)	208 (88.9%)	11 (5%)
Telephoned a physician or other health provider	232	1 (0%)	216 (93.1%)	15 (7%)

Table 8. Perceived effects of Internet use on health-related activities (experiences of telling health professionals about health-related information from the Internet) among Internet users

	n	Have done	Tried, but never done	Never tried
Have told health professionals about health-related information from the Internet	236	39 (17%)	12 (5%)	185 (78.4%)

Discussion

Principal Results

This study revealed four principal findings. First, the prevalence of Internet use via personal computer for acquiring health-related information was about one-quarter among those surveyed (23.8%), whereas the prevalence of Internet use via cell phone for this purpose was low (6%). The prevalence of Internet use via personal computer was higher than radio (19.1%), but lower than television (60.1%), newspapers (50.3%), and magazines (34.2%). Second, younger people, people with higher education levels, and people with higher household incomes were more likely to acquire health-related information by accessing the Internet via personal computer. Third, the prevalence of Internet use for health-related communication with health professionals, family, friends, or peers was small. Although cell phones were rarely used for this type of communication in general, 12.3% of respondents used cell phones for contacting family or friends specifically for health-related purposes. Finally, the majority of those using the Internet for health care purposes thought the Internet improved their health-related knowledge and affected their lifestyle attitudes, and felt confident after obtaining health-related information through the Internet. In contrast, less than one-quarter of respondents thought Internet use improved their ability to manage their health or changed their health-related activities. We further discuss these four findings below.

First, we found that the prevalence of Internet use via personal computer for health-related information was lower in Japan (24% in 2007) than in the United States (40% in 2001) [2] and Europe (42% in 2005 and 52% in 2007) [5,13]. On the other hand, the prevalence of using traditional sources of information in Japan, such as television and newspapers, was similar to that in the United States and Europe [3,5]. Although the Internet is increasingly being used as a source of health information [10], consumers still value and use traditional information sources in the United States and Europe [14,15,22]. Therefore, the Japanese general population may also still value traditional sources and not widely use the Internet to obtain health-related information.

Second, our results regarding characteristics of Internet users were consistent with many preceding studies pointing out that

XSL•FO
RenderX

older people, people with lower education levels, and people with lower household incomes reported less frequent access to the Internet [23-28]. Since these people may be unfamiliar with the Internet, these characteristics could result in a digital divide, a barrier to accessing health-related information through the Internet [29-33]. A generation gap in digital knowledge and skills is generally acknowledged [34]. Approximately 70% of people aged 50–64 use the Internet in the United States [26,27], whereas we found that hardly anyone over 50 years of age in Japan accessed the Internet. According to a white paper, the number of Internet users among the older people (over 65 years of age) has increased in recent years (28.1% in 2008, 36.9% in 2009) [35]. It is also suggested that active seniors who actively use the Internet could encourage other seniors to use the Internet. Some studies have proposed that less healthy people moderately use the Internet for health-related information [2,36], although some studies show that people with chronic disease are less likely than healthy people to have access to the Internet [37]. Although less healthy people are more likely to ask health professionals about information they find online [38], people who use the Internet for health purposes are more health oriented than people who do not search the Internet [39]. Therefore, people who use the Internet for health purposes might include both less healthy people, who use the Internet for recovery, and more healthy or health-oriented people, who use it for prevention. The relationship between health status and Internet searching behavior remains controversial.

Third, our results suggest that online communication generally remains uncommon in Japan. For communication with family, friends, or peers, cell phones were more used than personal computers. Cell phones were not used as a tool to acquire information, but as a tool for communication by people of all income levels. This could be because even average people in Japan can have advanced cell phones, which are frequently used for email communication with family or friends. For communication with health professionals, the Internet was less used in Japan than in the Unites States [2].

One reason why online communication generally remains uncommon in Japan might be the lack of systems related to eHealth in Japan. In the reimbursement payment system in Japan, the cost of health professional communications with patients is not reimbursed. In the Japanese context of universal health insurance coverage, treatments covered by insurance are not performed together with treatments not covered by insurance. Most health professionals and medical organizations do not promote this communication. Moreover, the legal system pertaining to personal medical information protection in Japan is not fully developed with regard to eHealth. The Japan Internet Medical Association (JIMA) was founded in 1998 to establish a framework for Internet medical usage [40]. JIMA created the Japanese version of the eHealth code of ethics [41], and has also developed the JIMA trust program. However, only 14 medical organizations obtained the JIMA trust mark, possibly because these ethics codes are self-imposed. Although the Ministry of Health, Labour and Welfare (MHLW) instituted the Guidelines on Security Management for Health Information Systems (first in 2005, fourth in 2010), there is no Act similar to the Health Insurance Portability and Accountability Act or Health Information Technology for Economic and Clinical Health Act in the United States [42]. Health professionals and medical organizations autonomously address privacy and security concerns associated with the electronic transmission of health information. Therefore, a nationwide privacy and security framework for eHealth is required in Japan.

The other reason why online communication generally remains uncommon in Japan could be the absence of a well-developed collaborative relationship among industry, government, and academia in Japan. In the United States, the vast majority of active eHealth services, such as WebMD, have been created by ventures put forth by cooperation and innovation among practitioners, researchers, and private industry [43]. Therefore, more collaborative efforts will be required in Japan. Some websites are being developed in Japan. Medical Information Network Distribution Service, which is operated by the Japan Council for Quality Health Care and funded by MHLW, has provided clinical practice guidelines in Japan on the Internet since 2004 [44-47]. Since the Medical Function Information Providing System was instituted by MHLW in 2007, prefectural governments have obligatorily provided information about the structure and outcomes of medical organizations on the Internet [47,48]. Websites created by nonprofit organizations and private industries, as well as pharmaceutical companies, have received awards for being the most informative health care websites in 2010 [49].

Fourth, our study showed that people tended to use the Internet for obtaining health-related information and felt confident in the information they obtained, which is compatible with many studies [36,50-52]. Obtaining information from the Internet, although it did not apparently change their activities, may encourage users to be confident that their ideas are supported. Nevertheless, our study also showed that few Internet users (6.6% via personal computer and 3.0% via cell phone) used the information for communication with health professionals. The frequency of patient communication with health professionals via the Internet was much lower than the frequency of patient visits to a physician's office (30.7% in a month) [53]. According to Hesse et al, people tend to go to the Internet first [11] and rarely share the information from the Internet with physicians [3]. They still trust face-to-face contact with physicians as their preferred source of health-related information [11,24,54]. The behavioral discrepancy between searching for information on the Internet and not using this information with health professionals might be due to user trust in health professionals, or to user conflict derived from untrusting health professionals whose attitude and behavior are incompatible with the information from the Internet. This is an important topic regarding communication between health providers and health consumers that should be addressed in the future.

Public Health Implications

Our findings have public health implications. Our results showed that Internet use of health-related information remains less common in Japan than in other developed countries [28]. Japanese aged 50–64 years, a large segment of the baby boomer generation that is going to require increased access to hospitals, did not access information on the Internet because of the digital

XSL•FO
RenderX

divide. As active seniors might lead other seniors to follow their example, it could be important to determine the needs of active seniors regarding Internet use. Our results also suggest a behavioral discrepancy. Inadequate use of information obtained from the Internet might have harmful consequences, such as Internet addiction [55,56] or cyberchondria, which is excessive health anxiety generated from online health searches [57,58]. To address these issues, we believe it is important to improve users' so-called eHealth literacy, defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [48,59,60]. It is appropriate to use the Internet as a supplement to health services rather than as a replacement [13,61-63], and to share the information with health professionals. There are differences between physicians and patients in health literacy [64], but it is also important that health professionals be mindful of patients' desire for health information [65] and the Internet presence [52,66]. They should discuss the information offered by patients and guide them to reliable and accurate health websites [52]. For searching websites, standards for eHealth, such as the e-Health Code of Ethics 2.0 [41], could be beneficial for both patients and health professionals. Health professionals, public health professionals, and eHealth developers should work together to educate patients about acquiring health information online and critically appraising it [67-69], and to provide tools for them to navigate to the highest-quality information [38].

Limitations

This study had some limitations. We acknowledge that the study's sample size was too small to examine the details of individuals who access the Internet via cell phone. The prevalence of Internet use via cell phone was lower than we had expected. Since this study aimed to measure the prevalence of Internet use for health-related information among the general Japanese population, a further study targeting the subset of Internet users who access the Internet via cell phone is required. We also acknowledge that there are no data about the response rate of respondents. In order to examine the extent of selection bias, we compared some indicative items of this survey with a national representative survey [16]. The distributions of age and sex in this survey were almost equal to those reported for the general Japanese population (see Multimedia Appendix 1 and Multimedia Appendix 2). The proportion of respondents using the Internet more than once a week was 41.5% in this survey and 48.5% in the national representative survey, calculated from data that 70.3% of Internet users (69.0% of respondents) use the Internet more than once a week. The small discrepancy between the national results and our findings can be attributed to the difference in survey methods; the national survey was conducted by mail. Nevertheless, given that the discrepancy was small, our results imply that the respondents to this survey were quasi-representative of the Japanese population. We accept that it is hard to discuss the accuracy of the prevalence of Internet use in Japan since there are no Japanese studies or data for comparison. Although repeated cross-sectional surveys are necessary to determine trends and associations, this study is useful in providing fundamental data in Japan.

Conclusions

In 2007, Japanese moderately used the Internet via personal computers for health purposes, and rarely used the Internet via cell phones. Older people, people with lower education levels, and people with lower household incomes were less likely to access the Internet via cell phone. The Internet moderately improved user health-related knowledge and attitudes, and encouraged user confidence in health-related information. However, it seldom changed their health-related abilities and activities, and was not often used for communicating with physicians. The paucity of Internet use for communication with physicians might be due to the payment system in Japan. Moreover, Internet users did not generally share the information they obtained from the Internet with health professionals. The health-related information from the Internet was inadequately used. Although cell phones were used as a communication tool for health purposes, the reimbursement payment system in Japan might be an obstacle to communication between health providers and health consumers. To encourage this communication, it is important to improve eHealth literacy, especially in middle-aged people. It is also important to make adequate amendments to the reimbursement payment system and nationwide eHealth privacy and security framework, and to develop a collaborative relationship among industry, government, and academia.

Acknowledgments

We are grateful to Professor Laurence C Baker, who provided the original questionnaire that was used in his work [2,29,36]. We also thank Mio Nakamura at Nippon Research Center Ltd for coordinating the survey. This survey was supported by a Grant-in-Aid for Scientific Research (A) from the Ministry of Education, Culture, Sports, Science and Technology, Japan to T Nakayama. The sponsor played no role in study design, data collection, analysis, or interpretation; in the writing of the paper; or in the decision to submit the paper for publication.

Conflicts of Interest

None declared

Authors' Contributions

YT designed the protocol, was responsible for the data analysis, and drafted the manuscript. TO, SO, and TN gave valuable advice for developing the questionnaire. TO, TI, and TN supported drafting the manuscript and analyzing the data. All authors gave valuable advice for designing the protocol and preparing the manuscript. TN conceived the study and acted as supervisor.

XSL•FO
RenderX

Multimedia Appendix 1

Characteristics of survey participants and the Japanese population.

[PDF file (Adobe PDF File), 52 KB - jmir_v13i4e110_app1.pdf]

Multimedia Appendix 2

Health status of survey participants and the Japanese population.

[PDF file (Adobe PDF File), 61 KB - jmir_v13i4e110_app2.pdf]

References

1. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
2. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406 [FREE Full text] [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
3. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624 [FREE Full text] [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
4. Ybarra ML, Suman M. Help seeking behavior and the Internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]
5. Kummervold PE, Chronaki CE, Lausen B, Prokosch HU, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
6. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
7. Takahashi Y, Uchida C, Miyaki K, Sakai M, Shimbo T, Nakayama T. Potential benefits and harms of a peer support social network service on the internet for people with depressive tendencies: qualitative content analysis and social network analysis. J Med Internet Res 2009;11(3):e29 [FREE Full text] [doi: 10.2196/jmir.1142] [Medline: 19632979]
8. Van De Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of Health 2.0 and Medicine 2.0: a systematic review. J Med Internet Res 2010;12(2):e18 [FREE Full text] [doi: 10.2196/jmir.1350] [Medline: 20542857]
9. National Cancer Institute. Health Information National Trends Survey URL: http://hints.cancer.gov/ [accessed 2011-01-04] [WebCite Cache ID 5vSaRgIle]
10. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
11. Hesse BW, Moser RP, Rutten LJ. Surveys of physicians and electronic health information. N Engl J Med 2010 Mar 4;362(9):859-860 [FREE Full text] [doi: 10.1056/NEJMc0909595] [Medline: 20200398]
12. Sorensen T. Norwegian Centre for Telemedicine. WHO/European eHealth Consumer Trends Survey URL: http://www.telemed.no/index.php?id=275869 [accessed 2011-01-04] [WebCite Cache ID 5vSamiDR9]
13. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
14. Warner D, Procaccino JD. Women seeking health information: distinguishing the web user. J Health Commun 2007 Dec;12(8):787-814. [doi: 10.1080/10810730701672090] [Medline: 18030642]
15. Dumitru RC, Bürkle T, Potapov S, Lausen B, Wiese B, Prokosch HU. Use and perception of internet for health related purposes in Germany: results of a national survey. Int J Public Health 2007;52(5):275-285. [Medline: 18030943]
16. Ministry of Internal Affairs and Communications (Japan). 2008 Apr 18. Communications Usage Trend Survey in 2007 Compiled URL: http://www.soumu.go.jp/johotsusintokei/tsusin_riyou/data/eng_tsusin_riyou02_2007.pdf [accessed 2011-01-04] [WebCite Cache ID 5vSnyDV69]
17. Tabuchi H. The New York Times. 2009 Jul 19. Why Japan's Cellphones Haven't Gone Global URL: http://www.nytimes.com/2009/07/20/technology/20cell.html?_r=1 [accessed 2011-01-03] [WebCite Cache ID 5vSb1FveN]
18. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
19. Giustini D. How Web 2.0 is changing medicine. BMJ 2006 Dec 23;333(7582):1283-1284. [doi: 10.1136/bmj.39062.555405.80] [Medline: 17185707]
20. Anonymous. Ministry of Internal Affairs and Communications (Japan). 2008. White Paper 2008: Information and Communications in Japan URL: http://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2008/2008-index.html [WebCite Cache ID 5vSbCAid]

XSL•FO
RenderX

21. Nippon Research Center Ltd. 2006. URL: http://www.nrc.co.jp/english/index.html [accessed 2011-01-03] [WebCite Cache ID 5vSbKE1u5]
22. Dumitru RC, Ganslandt T, Prokosch HU. German healthcare consumer's perception of the Internet as a source of health related information. AMIA Annu Symp Proc 2006:224-228. [Medline: 17238336]
23. Dickerson S, Reinhart AM, Feeley TH, Bidani R, Rich E, Garg VK, et al. Patient Internet use for health information at three urban primary care clinics. J Am Med Inform Assoc 2004;11(6):499-504 [FREE Full text] [doi: 10.1197/jamia.M1460] [Medline: 15298993]
24. Välimäki M, Nenonen H, Koivunen M, Suhonen R. Patients' perceptions of Internet usage and their opportunity to obtain health information. Med Inform Internet Med 2007 Dec;32(4):305-314. [doi: 10.1080/14639230701819792] [Medline: 18072007]
25. Miller EA, West DM. Characteristics associated with use of public and private web sites as sources of health care information: results from a national survey. Med Care 2007 Mar;45(3):245-251. [doi: 10.1097/01.mlr.0000244509.60556.49] [Medline: 17304082]
26. Kaiser Family Foundation. 2005. e-Health and the Elderly: How Seniors Use the Internet for Health - Survey URL: http://www.kff.org/entmedia/upload/e-Health-and-the-Elderly-How-Seniors-Use-the-Internet-for-Health-Information-Key-Findings-From-a-National-Survey-of-Older-Americans-Survey-Report.pdf [accessed 2011-01-03] [WebCite Cache ID 63FjzcuqA]
27. Voelker R. Seniors seeking health information need help crossing "digital divide". JAMA 2005 Mar 16;293(11):1310-1312. [doi: 10.1001/jama.293.11.1310] [Medline: 15769951]
28. Weaver JB, Mays D, Lindner G, Eroglu D, Fridinger F, Bernhardt JM. Profiling characteristics of internet medical information users. J Am Med Inform Assoc 2009;16(5):714-722 [FREE Full text] [doi: 10.1197/jamia.M3150] [Medline: 19567794]
29. Wagner TH, Bundorf MK, Singer SJ, Baker LC. Free internet access, the digital divide, and health information. Med Care 2005 Apr;43(4):415-420. [Medline: 15778645]
30. Brodie M, Flournoy RE, Altman DE, Blendon RJ, Benson JM, Rosenbaum MD. Health information, the Internet, and the digital divide. Health Aff (Millwood) 2000;19(6):255-265 [FREE Full text] [Medline: 11192412]
31. Hsu LL. An exploratory study of Taiwanese consumers' experiences of using health-related websites. J Nurs Res 2005 Jun;13(2):129-140. [Medline: 15986314]
32. Rains SA. Perceptions of traditional information sources and use of the world wide web to seek health information: findings from the health information national trends survey. J Health Commun 2007;12(7):667-680. [doi: 10.1080/10810730701619992] [Medline: 17934943]
33. Gracia E, Herrero J. Internet use and self-rated health among older people: a national survey. J Med Internet Res 2009;11(4):e49 [FREE Full text] [doi: 10.2196/jmir.1311] [Medline: 19955041]
34. DeMaria AN. The digital divide. J Am Coll Cardiol 2008 Feb 19;51(7):771-772. [doi: 10.1016/j.jacc.2008.01.009] [Medline: 18279743]
35. Ministry of Internal Affairs and Communications (Japan). 2010. 2010 White Paper: Information and Communications in Japan URL: http://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2010/2010-index.html [accessed 2011-07-18] [WebCite Cache ID 60GWZmRA9]
36. Wagner TH, Baker LC, Bundorf MK, Singer S. Use of the Internet for health information by the chronically ill. Prev Chronic Dis 2004 Oct;1(4):A13 [FREE Full text] [Medline: 15670445]
37. Fox S, Purcell K. Chronic disease and the Internet. Washington, DC: Pew Internet & American Life Project; 2010 Mar 24. URL: http://www.pewinternet.org/~/media//Files/Reports/2010/PIP_Chronic_Disease_with_topline.pdf [accessed 2011-11-08] [WebCite Cache ID 632sjyGez]
38. Houston TK, Allison JJ. Users of Internet health information: differences by health status. J Med Internet Res 2002;4(2):E7 [FREE Full text] [doi: 10.2196/jmir.4.2.e7] [Medline: 12554554]
39. Dutta-Bergman M. Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. J Med Internet Res 2003;5(3):e21 [FREE Full text] [doi: 10.2196/jmir.5.3.e21] [Medline: 14517112]
40. Tatsumi H, Mitani H, Haruki Y, Ogushi Y. Internet medical usage in Japan: current situation and issues. J Med Internet Res 2001;3(1):E12 [FREE Full text] [doi: 10.2196/jmir.3.1.e12] [Medline: 11720954]
41. Japan Internet Medical Association. 2007. e-Health Code of Ethics 2. 0 URL: http://www.jima.or.jp/ehealth_code/JIMAeHealth_code20(English).pdf [accessed 2011-01-03] [WebCite Cache ID 5vSbzkkx8]
42. Buntin MB, Jain SH, Blumenthal D. Health information technology: laying the infrastructure for national health reform. Health Aff (Millwood) 2010 Jun;29(6):1214-1219. [doi: 10.1377/hlthaff.2010.0503] [Medline: 20530358]
43. WebMD.com. URL: http://www.webmd.com/ [accessed 2011-07-18] [WebCite Cache ID 60GWffWUS]
44. Japan Council for Quality Health Care. Medical Information Network Distribution Service (Minds) (in Japanese). 2005. URL: http://minds.jcqhc.or.jp/ [accessed 2011-07-18] [WebCite Cache ID 60GWlLXEY]
45. Sato Y, Nakayama T, Satoh T, Hoshi K, Kojimahara N, Miyaki K. Preferred information media for providing clinical practice guidelines to physicians in Japan: a needs assessment study by the Medical Information Network Distribution Service (Minds). Gen Med 2006;7(2):45-52.

XSL•FO
RenderX

46. Nakayama T. What are "clinical practice guidelines"? J Neurol 2007;254(Suppl 5):2-7. [doi: 10.1007/s00415-007-5001-y]
47. Ministry of Health, Labour and Welfare (Japan). Medical Function Information Providing System (in Japanese) URL: http://www.mhlw.go.jp/topics/bukyoku/isei/teikyouseido/ [accessed 2011-07-18] [WebCite Cache ID 60GWseiat]
48. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. [Developing Japanese version of the eHealth Literacy Scale (eHEALS)]. Nihon Koshu Eisei Zasshi 2011 May;58(5):361-371. [Medline: 21905612]
49. QLife, Inc. 2010. The Award of the Best Enlightenment Healthcare Website (in Japanese) URL: http://www.qlife.jp/award/ [accessed 2011-07-18] [WebCite Cache ID 60GX4Hf2j]
50. Beckjord EB, Finney Rutten LJ, Squiers L, Arora NK, Volckmann L, Moser RP, et al. Use of the internet to communicate with health care providers in the United States: estimates from the 2003 and 2005 Health Information National Trends Surveys (HINTS). J Med Internet Res 2007;9(3):e20 [FREE Full text] [doi: 10.2196/jmir.9.3.e20] [Medline: 17627929]
51. Hart A, Henwood F, Wyatt S. The role of the Internet in patient-practitioner relationships: findings from a qualitative research study. J Med Internet Res 2004 Sep 30;6(3):e36 [FREE Full text] [doi: 10.2196/jmir.6.3.e36] [Medline: 15471762]
52. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
53. Fukui T, Rhaman M, Takahashi O, Saito M, Shimo T, Endo H. The ecology of medical care in Japan. Jpn Med Assoc J 2005;48:163-167.
54. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188. [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
55. Miller G, Holden C. Psychiatry. Proposed revisions to psychiatry's canon unveiled. Science 2010 Feb 12;327(5967):770-771. [doi: 10.1126/science.327.5967.770-a] [Medline: 20150459]
56. Young K. Internet addiction: diagnosis and treatment considerations. J Contemp Psychother 2009;39:241-246. [doi: 10.1007/s10879-009-9120-x]
57. White R, Horvitz E. Cyberchondria: studies of the escalation of medical concerns in Web search. ACM Trans Inf Syst 2009;27(4):23. [doi: 10.1145/1629096.1629101]
58. Harding KJ, Skritskaya N, Doherty E, Fallon BA. Advances in understanding illness anxiety. Curr Psychiatry Rep 2008 Aug;10(4):311-317. [Medline: 18627669]
59. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
60. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
61. Bouche G, Migeot V. Parental use of the Internet to seek health information and primary care utilisation for their child: a cross-sectional study. BMC Public Health 2008;8:300 [FREE Full text] [doi: 10.1186/1471-2458-8-300] [Medline: 18755029]
62. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
63. Lee CJ. Does the internet displace health professionals? J Health Commun 2008;13(5):450-464. [doi: 10.1080/10810730802198839] [Medline: 18661387]
64. Safeer RS, Keenan J. Health literacy: the gap between physicians and patients. Am Fam Physician 2005 Aug 1;72(3):463-468 [FREE Full text] [Medline: 16100861]
65. Xie B. Older adults' health information wants in the internet age: implications for patient-provider relationships. J Health Commun 2009 Sep;14(6):510-524. [doi: 10.1080/10810730903089614] [Medline: 19731124]
66. Gorrindo T, Groves JE. Web searching for information about physicians. JAMA 2008 Jul 9;300(2):213-215. [doi: 10.1001/jama.2008.44] [Medline: 18612120]
67. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734 [FREE Full text] [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
68. Peterson G, Aslani P, Williams KA. How do consumers search for and appraise information on medicines on the Internet? A qualitative study using focus groups. J Med Internet Res 2003 Dec 19;5(4):e33 [FREE Full text] [doi: 10.2196/jmir.5.4.e33] [Medline: 14713661]
69. Takahashi Y, Sakai M, Fukui T, Shimbo T. Measuring the ability to interpret medical information among the Japanese public and the relationship with inappropriate purchasing attitudes of health-related goods. Asia Pac J Public Health 2011 May;23(3):386-398. [doi: 10.1177/1010539509344882] [Medline: 19723690]

Abbreviations

JIMA: Japan Internet Medical Association
MHLW: Ministry of Health, Labour and Welfare

XSL•FO
RenderX

Edited by G Eysenbach; submitted 09.03.11; peer-reviewed by T van Mierlo, S Lim Choi Keung, YS Bin; comments to author 05.05.11; revised version received 27.07.11; accepted 26.08.11; published 14.12.11.

Please cite as:
Takahashi Y, Ohura T, Ishizaki T, Okamoto S, Miki K, Naito M, Akamatsu R, Sugimori H, Yoshiike N, Miyaki K, Shimbo T, Nakayama T
Internet Use for Health-Related Information via Personal Computers and Cell Phones in Japan: A Cross-Sectional Population-Based Survey
J Med Internet Res 2011;13(4):e110
URL: http://www.jmir.org/2011/4/e110/
doi:10.2196/jmir.1796
PMID:22169526

©Yoshimitsu Takahashi, Tomoko Ohura, Tatsuro Ishizaki, Shigeru Okamoto, Kenji Miki, Mariko Naito, Rie Akamatsu, Hiroki Sugimori, Nobuo Yoshiike, Koichi Miyaki, Takuro Shimbo, Takeo Nakayama. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving Access to Information and Support for Patients With Less Common Cancers: Hematologic Cancer Patients' Views About Web-Based Approaches

Christine Louise Paul[1], BA(Hons), PhD; Mariko Leanne Carey[1], BSC(Hons), DPsych; Alix Edna Hall[1], BPsych(Hons); Marita Clare Lynagh[1], BHMSc(Hons), GradDip(HlthProm), PhD; Robert W Sanson-Fisher[1], BPsych(Hons), MPsych, PhD, DSc; Frans Alexander Henskens[1], BMath, Dip Ed, GradDip(CompSci), PhD

Health Behaviour Research Group, Priority Research Centre for Health Behaviour, Hunter Medical Research Institute, School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

Corresponding Author:
Christine Louise Paul, BA(Hons), PhD
Health Behaviour Research Group, Priority Research Centre for Health Behaviour, Hunter Medical Research Institute
School of Medicine and Public Health
University of Newcastle
Room 268, Level 2, David Maddison Building
University Drive
Callaghan, 2308
Australia
Phone: 61 2 49138472
Fax: 61 2 49138779
Email: chris.paul@newcastle.edu.au

Abstract

Background: Meeting the psychosocial needs of vulnerable groups such as cancer survivors remains an ongoing challenge. This is particularly so for those who have less access to the usual forms of medical specialist and in-person support networks. Internet-based approaches offer an opportunity to better meet patients' information and support needs by overcoming the barrier of geographic isolation.

Objective: The aim of the study was to assess the reported level of access to the Internet, preferred sources of information, and preferred sources of support among survivors of hematologic cancers.

Method: A population-based, Australian state cancer registry invited eligible survivors to complete a survey about psychosocial needs, including items measuring Internet access and patterns of use. Of the 732 eligible survivors invited to participate, 268 (36.6%) completed and returned the pen-and-paper-based survey.

Results: The majority of participants (186/254, 73.2%) reported a high level of access to the Internet, with higher Internet access associated with a higher level of education, larger household, younger age, and being married or employed. A total of 62.2% (156/251) of survivors indicated they were likely to use the Internet for accessing information, with the percentage much lower (69/251, 28%) for accessing support via the Internet. Likelihood of using the Internet for support was associated with feeling anxious and being employed.

Conclusions: While the Internet appears to offer promise in increasing equitable access to information and support for cancer survivors for both metropolitan and regional areas, it is viewed less favorably for support and by particular subgroups (eg, older people and those without a university degree) within the survivor population. Promoting greater understanding of this mode of support may be required to achieve its potential. Information and support options other than Web-based approaches may continue to be needed by vulnerable groups of cancer survivors.

(J Med Internet Res 2011;13(4):e112) doi:10.2196/jmir.1894

KEYWORDS

Cancer; hematologic diseases; information-seeking behavior; social support

XSL•FO
RenderX

Introduction

Although hematologic cancers such as lymphoma and leukemia are much less prevalent than other cancer types such as colorectal cancer or breast cancer [1], in developed countries such as Australia they are a major cause of cancer death, due to poor survival rates compared with other cancers [1]. Diagnosis and treatment can have a devastating impact on life expectancy, fertility and sexuality [2,3], and overall health [4]. Accordingly, these patients report a need for information [5] and support [6]. Canadian data indicate that rates of clinical distress among those with hematologic cancers can range from 32% to 48% given the debilitating nature of the disease and its treatment [7]. Australian data indicate that hematologic cancer patients are also often isolated from support systems due to the need to travel to major hospitals for treatment, with treatment potentially lasting several months [5,8].

The prevalence and nature of the disease and its treatment raise some particular issues for the provision of psychosocial support. The opportunities for peer support, for example, are limited by the relative rarity of an age- or gender-matched survivor being available for either face-to-face or telephone-based peer-support programs. The concentration of hematologic professionals in major centers also can result in low access to face-to-face information and support once a patient has completed a round of treatment, particularly for those who live in nonmetropolitan locations. Further, opportunities for social and peer support may be limited due to lengthy inpatient stays and restriction of activities due to risk of neutropenic infection. Therefore, it is likely that a suite of options needs to be made available in order that hematologic cancer patients receive sufficient information and support throughout the months and years that may follow diagnosis.

Alongside the vital role of specialist medical staff, the Internet offers unique advantages for the delivery of information and psychosocial support to hematologic cancer patients, primarily due to its high level of accessibility. Up to 77% of Australian cancer patients access information about cancer via the Internet [9]. Internet access in Australia has quadrupled between 1998 and 2008 [10]. The most recent Australian data suggest that 72% of the population have home Internet access [10], while in the United States up to 69% of people may have home Internet access [11]. For those in regional and remote areas [12] the Internet may overcome some geographic barriers. It provides a way of connecting with information, services, and others in a similar situation no matter their location or level of wellness. It also offers the opportunity to provide peer online forums to obtain support from others in similar positions, who may not be accessible face-to-face.

A small group of studies have explored the effectiveness of Web-based psychosocial support for cancer survivors using robust randomized controlled designs [13-17], with mixed findings for psychosocial outcomes. The single study that included hematologic cancer survivors [15] also involved participants with other types of cancer diagnoses and suggested that those who were single, older, and less educated were less likely to use the Web-based intervention. Issues of reach and access in relation to Web-based interventions have also rarely been addressed.

While Internet accessibility is apparently high and increasing, there are no current data about the accessibility of this resource for hematologic cancer patients. Internet access can differ according to income, education, age, and geographic location [12,18]. These differences may in turn create or exacerbate inequality. Given that the sociodemographic profile of adult hematologic cancer patients includes a substantial proportion of older age groups [1], it is important to establish whether older or disadvantaged patients have ready access to the Internet in a manner that is conducive to its use for obtaining support and information.

The study aimed to do the following in a cross-sectional sample of people with a diagnosis of a hematologic cancer: (1) investigate the proportion of metropolitan versus regional survivors who reported a high level of access to the Internet, (2) measure the proportion who reported being likely to use various sources (Internet, print, telephone, face-to-face) for information and support and the perceived benefits of Internet options, and (3) explore the sociodemographic characteristics of survivors who reported both a high level of Internet access and being likely to use the Internet for information or support.

Methods

Design

Sample

Through a population-based cancer registry we recruited survivors aged 18 to 80 years at study invitation who had a diagnosis of leukemia, lymphoma, or myeloma in the prior 3 years. Use of this registry permitted sampling across the full range of cancer types, locations, and stages of treatment.

Procedure

On behalf of the researchers the cancer registry sent all eligible patients a questionnaire package containing an invitation letter, information statement, prepaid envelope, registry brochure, self-report pen-and-paper survey, and questionnaire package for their principal support person. Patients who did not respond to the initial questionnaire after 4 weeks were mailed a reminder letter from the cancer registry and a second questionnaire package.

Measure

The 30-minute self-report pen-and-paper survey comprised a series of measures regarding psychosocial issues for cancer survivors, a subset of which are reported here. Participants were asked about their use of the Internet, accessibility of the Internet, likelihood of using each of a range of options for seeking support or assistance, and perceived benefits and disadvantages of the Internet for cancer-related information and support. Multimedia Appendix 1 contains the Internet-related survey items. The Depression Anxiety Stress Scales [19], a reliable and valid measure for assessing psychological status in cancer patients [20], was also completed as part of the survey. Diagnosis, gender, age, and postcode (to assess metropolitan status) were obtained from registry records with the patient's permission.

XSL•FO
RenderX

Analysis

Metropolitan Versus Regional Categorization

Survivors' residential postcodes were used to classify their location on the Accessibility/Remoteness Index of Australia (ARIA+) classification. Metropolitan was defined as the ARIA+ category major cities, and regional was defined as inner regional, outer regional, remote, or very remote.

Level of Internet Access

We reported proportions to describe level of Internet access on each access item. Chi-squares were used to compare metropolitan versus regional access on each item and on overall access score. An access score was calculated as follows. A high score consisted of 5 or more of the following responses: frequency of access (any/most of the time), connection problems (none/minor), privacy (moderately/very), comfort (very/moderately), printing (any/limited), and confidence (very/moderately). A moderate score was any 3 or 4 of these responses, and low was classed as a score of 0–2. A score of 0 was given to those who indicated they had no access to the Internet for personal use.

Likelihood of Using Various Modes of Information and Support

Response categories of likely and very likely were combined. Proportions and 95% confidence intervals were used to describe the data for each item.

Sociodemographic Characteristics Associated With Reported Internet Access and Likelihood of Using the Internet

We conducted initial chi-square analyses with the following independent variables: gender, living in a rural area, education, marital status, employment status, household size, health status, and whether the survivor had normal or some level of anxiety or depression. Age at diagnosis in 5-year categories was analysed using *t* tests. The dependent variables were Internet access (high access versus low/no access), and the likelihood of using the Internet as a source of each of information and support (likely/very likely compared with unsure/not likely/very unlikely). Those independent variables with a $P < .25$ were included in a backward stepwise logistic regression for each dependent variable. We removed variables until we found an optimal model, based on the Bayesian information criterion. Analyses were conducted in Stata version 11.1 (StatCorp LP, College Station, TX, USA).

Results

Sample

We invited 732 eligible survivors to complete and return a survey. Of these, 268 (36.6%) returned a completed survey. The age distribution of responders was significantly different from that of nonresponders, with younger people less likely to return a survey than older people ($\chi^2_5 = 17.2$, $P = .004$). Gender, area of residence, type of cancer, and year of diagnosis were not significantly different between responders and nonresponders. As Table 1 shows, participants from a regional location were significantly older and less likely to be employed than those from metropolitan locations. There were no differences between regional and metropolitan participants in terms of cancer type, gender, education, and marital status (see Table 1).

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of the sample calculated for those living in a major city or regional area at the time of the survey (n = 268)

	Metropolitan		Regional		Total		Test	*P* value
	n	%	n	%	n	%		
Age (years) (Mean, SD)	(57.4, 14.3)		(61.9, 12.0)		(59.5, 13.4)		$F_{1,237} = 6.74$	.01
Female	60	40%	51	43%	111	41.4%	$\chi^2_1 = 0.2$	.67
Cancer type								
Lymphoma	13	9%	7	6%	20	8%		
Leukemia	43	29%	29	24%	72	27%		
Myeloma	25	17%	17	14%	42	16%		
Non-Hodgkin lymphoma	68	46%	66	56%	134	50.0%	$\chi^2_3 = 2.8$	.43
Education [a]								
High school or less	60	40%	50	42%	110	41.0%		
Vocational training	55	37%	48	40%	103	38.4%		
University	33	22%	20	17%	53	20%	$\chi^2_2 = 1.2$	.55
Employed	81	54%	41	34%	122	45.5%	$\chi^2_1 = 11.3$	.001
Married	108	72.5%	96	81%	204	76.1%	$\chi^2_1 = 1.7$	.20
Total	149	55.6%	119	44.4%	268			

[a] Education data were missing for two participants.

Level of Internet Access

Of the 260 participants who answered the Internet access questions, 204 (78.5%) reported having home Internet access and 67 (26%) reported Internet access at work; 48 (19%) reported having no Internet access and a further 5 (2%) reported no access to the Internet for personal use—that is, 20% were without access to the Internet for personal use. Of those with access (n=207), 167 (80.7%) report daily or weekly use of email.

Table 2 describes the nature of reported Internet access, indicating that approximately 73% of participants reported high levels of Internet access, with regional participants more likely to report connection problems.

XSL•FO
RenderX

Table 2. Nature of Internet access for those with access who answered all the access questions (n = 201), and overall level of access for whole sample (n = 254)

Nature of access	Metropolitan (n=111)		Regional (n=90)		Total		χ^2_2	*P* value
	n	%	n	%	n	%		
Frequency of access								
Any time	97	87%	74	82%	171	85.1%		
Most of time	13	12%	13	14%	26	13%	1.92	.38
Connection problems								
None	90	81%	56	62%	146	73%		
Minor	20	18%	31	34%	51	25%	9.20	.01
Private								
Very	66	60%	66	73%	132	65.7%		
Moderately	40	36%	19	21%	59	29%	5.34	.07
Comfortable								
Very	84	76%	67	74%	151	75.1%		
Moderately	27	24%	22	24%	49	24%	1.24	.54
Can print personal information								
Any amount	98	88%	73	81%	171	85.1%		
Limited amount	6	5%	8	9%	14	7%	2.02	.36
Confident with Internet								
Very	60	54%	42	47%	102	50.8%		
Moderately	41	37%	29	32%	70	35%	5.90	.05
Access score[a]								
High	106	75.7%	80	70%	186	73.2%		
Moderate	5	4%	10	9%	15	6%		
Low	0	0%	0	0%	0	0%		
None	29	21%	24	21%	53	21%	3.14	.21

[a] See text for access score calculation. The denominator for access score is the whole sample (ie, includes those with no access).

Likelihood of Using Various Modes of Information and Support

As Table 3 shows, face-to-face and print were the preferred approaches for receiving both information and support. Approximately 62% of the sample reported they were likely to use the Internet for information, while 27% reported being likely to use the Internet to access support. The main perceived benefits of use of the Internet as a source of either information or support were that it is available anytime (137/253 = 54.2%) and contains a large amount of information (105/253 = 41.5%). A minority of respondents mentioned additional benefits of not needing to travel (81/253 = 32%), low cost (79/253 = 31%), and not requiring personal contact (44/253 = 17%). The perceived disadvantages of Internet-based support were a lack of specificity (102/251 = 40.6%), being too complex (85/251 = 34%), being too impersonal (69/251 = 28%), and difficulty with using the Internet (35/251 = 14%).

XSL•FO
RenderX

Table 3. Likelihood (likely/very likely) of using Internet, telephone, print, electronic media, or face-to-face forms of support (n = 251)

Mode	Use for information		Use for support	
	n	% (95% CI[a])	n	% (95% CI[a])
Face-to-face	218	87% (83%–91%)	209	83% (79%–88%)
Print	204	81% (76%–86%)	164	65% (59%–71%)
Internet	156	62% (56%–68%)	69	27% (22%–33%)
Electronic	137	55% (48%–61%)	107	43% (36%–49%)
Telephone	131	52% (46%–58%)	96	38% (32%–44%)
Number of options chosen as likely or very likely				
≥2	222	88% (84%–92%)	170	68% (62%–74%)
1 only	21	8% (5%–12%)	65	26% (20%–31%)
Print only[b]	4	2% (0%–3%)	11	4% (2%–7%)
Face-to-face only[b]	13	5 (2%–8%)	52	21% (16%–26%)
None	8	3 (1%–5%)	16	6% (3%–9%)

[a] Confidence interval.

[b] Likely/very likely for item of interest and unsure/not likely/very unlikely to all others.

Sociodemographic Characteristics Associated With Reported Internet Access and Likelihood of Use

Educational level was significantly associated with reported high Internet access (Fisher exact test $P < .001$) but was not included in the multiple logistic regression model due to a zero cell count (all 53 university-educated participants reported high Internet access). Household size was also associated with high Internet access, with 74% (90/121) of those living with 1 other person and 97% (55/57) of those living with 2 or more people having high access, compared with 66% (21/32) of those living alone (Fisher exact test $P < .001$). This was also not included in the model due to low cell counts. The multiple logistic regression (see Table 4) indicated that younger people were more likely than those who were older to report high Internet access, as were those who were married and those in full- or part-time employment compared respectively with those who were single or not employed. Those who reported that they were likely to use the Internet to find information were more likely to be younger rather than older, to be anxious rather than not anxious, and to have a university degree than were those with only a high school education or vocational training. Participants who were feeling anxious and those in full- or part-time employment, compared with those not employed, were more likely to report being likely to use the Internet as a means of support.

XSL•FO
RenderX

Table 4. Logistic regression analysis of factors associated with high reported Internet access, likelihood of using the Internet for information, and likelihood of using the Internet for support

	High or likely	Low or unlikely	Odds ratio (95% CI[a])	*P* value
High access (n = 210)				
Age (5 years) , mean (SD)	52.7 (13.2)	65.46 (7.6)	0.89 (0.84–0.94)	<.001
Married , n (%)				
No	30 (65%)	16 (35%)		
Yes	136 (82.9%)	28 (17%)	5.63 (2.12–14.94)	.001
Employed, n (%)				
No	69 (65%)	38 (36%)		
Yes	97 (94%)	6 (6%)	4.02 (1.37–11.8)	.01
Likely to use the Internet for information (n = 221)				
Age (5 years), mean (SD)	52.12 (13.87)	60.68 (10.41)	0.95 (0.93–0.98)	<.001
Education, n (%)				
High school only	45 (51%)	43 (49%)		
Vocational training	50 (59%)	35 (41%)	1.3 (0.68–2.46)	.43
University degree	41 (85%)	7 (15%)	5.06 (1.97–12.98)	.001
Anxious, n (%)				
No	84 (56%)	66 (44%)		
Yes	52 (73%)	19 (27%)	2.39 (1.23–4.63)	.01
Likely to use the Internet for support (n = 221)				
Employed, n (%)				
No	22 (20%)	91 (81%)		
Yes	37 (34%)	71 (66%)	2.53 (1.33–4.81)	.005
Anxious, n (%)				
No	30 (20%)	120 (80.0%)		
Yes	29 (41%)	42 (59%)	3.17 (1.66–6.05)	<.001

[a] Confidence interval.

Discussion

As approximately three-quarters (73%) of the sample reported a high level of Internet access, such an approach appears to be relatively accessible to most patients. It must, however, be acknowledged that a substantial minority of the sample (20%) reported either having no Internet access at all (18%) or no access for personal use (2%). As higher reported Internet access was associated with higher educational level, younger age, being married, and being employed, those with less access appear to be a potentially isolated and disadvantaged group. Therefore, in order to avoid creating inequity, care should be taken to develop and provide appropriate alternative forms of information and support for such patients. An unexpected finding was that of no reported differences between regional and metropolitan participants regarding access to the Internet, other than greater difficulties with connectivity in regional areas. Therefore, Web-based approaches may indeed assist with improving access to information and support for cancer survivors, across a range of geographic locations. It is likely that adult patients with cancers other than hematologic cancers would similarly benefit from access to Web-based options for information and support.

The reported likelihood of using the Internet for obtaining information or seeking support was relatively low at 62% and 27%, respectively. Studies of Web-based interventions for depression and anxiety found that 78% to 95% of participants took up the offer [21-24]. It may be that the concrete offer of a Web-based program at a time of need is more engaging than the hypothetical possibility proposed in the present study. The samples of patients with a mental illness were younger than the hematologic cancer patient sample and, therefore, likely to be more familiar with Web-based technology.

The data suggest that, while the vast majority of patients reported being likely to use multiple modes for gathering information or seeking support (88% and 68%, respectively), face-to-face and print were the generally preferred forms. Notably, almost 1 in 4 participants reported they would access only one form of support, suggesting that retaining a range of support options is required in order to cater for the support needs of all hematologic cancer patients.

XSL•FO
RenderX

Interestingly, patients' preference for receiving information via face-to-face or print mode has not changed over time, despite increased accessibility to Internet resources. Previous work by Hinds et al suggested that cancer patients receiving radiotherapy preferred to receive verbal information from their physician in the pretreatment phase and printed information in the posttreatment phase [25]. A more recent systematic review that examined information needs and sources of information across a wider range of cancer patients found that the most frequently cited sources of information were health care professionals and printed materials [26].

In accordance with our findings, one other study has found that cancer patients who were single, older, male, and less educated [15] were less likely than their counterparts to engage with Web-based forms of information or support. Therefore, while Web-based provision of information may be attractive to the majority of patients, those who are less interested in such formats should not be forgotten. The perceived disadvantages of the Internet, particularly complexity and impersonality, also suggest the development of customized Web-based information sources may be useful for patients, rather than relying on generic engine-based searches such as Google. Promotion, careful training, and assistance may reduce some reticence toward newer forms of technology, although print or face-to-face options may need to be retained for those who continue to need or prefer such forms of communication.

An additional new finding is the association between being classified as anxious and a reported likelihood of using the Internet for information and support. This may reflect a greater need or desire for information and support among this group, potentially driving a desire to access available options. Alternatively, anxious cancer survivors may prefer options that require less interpersonal interaction. Other studies support the view that level of anxiety mediates the relationship between seeking information online and using health care services [27].

Limitations

The low response rate limits the generalizability of the data. However, given the scarcity of data regarding Internet accessibility for cancer patients generally and hematologic cancer patients in particular, these data are the best estimates available. It is possible that a paper-based survey is less likely to be completed by those with a preference for electronic media, resulting in the data providing an underestimate of respondents' likely use of the Internet as a source of information or support. Low rates of expected use of the Internet, particularly for support, may also be partly due to difficulties in conceptualizing how such support might operate.

Conclusions

Ensuring that all hematologic cancer patients have equitable access to information and support remains a challenge. While Web-based approaches to information provision appear likely to be accessible and acceptable to the majority of patients, they are less attractive for the provision of support. In addition, more vulnerable patients such as those who are older, single, unemployed, or less educated are particularly likely to require alternative forms of information and support.

Acknowledgments

The authors gratefully acknowledge the study funding received from Cancer Australia, *beyondblue*, *Psycho-oncology Cooperative Group*, and infrastructure support from the Hunter Medical Research Institute. The authors also greatly appreciate the efforts of the survey participants and Dr Emilie Cameron for data analysis.

Conflicts of Interest

None declared

Authors' Contributions

CP oversaw the writing of the manuscript. AH conducted the data collection. All authors contributed equally to survey development and sampling, and to writing and reviewing the manuscript.

Multimedia Appendix 1

Survey items relating to Internet access and likelihood of use.

[PDF file (Adobe PDF File), 87 KB - jmir_v13i4e112_app1.pdf]

References

1. Australian Institute of Health and Welfare, Australasian Association of Cancer Registries. Cancer in Australia: An Overview, 2010. Canberra: AIHW; 2010.
2. Leader A, Lishner M, Michaeli J, Revel A. Fertility considerations and preservation in haemato-oncology patients undergoing treatment. Br J Haematol 2011 May;153(3):291-308. [doi: 10.1111/j.1365-2141.2011.08629.x] [Medline: 21391973]
3. Schover LR. Sexuality and fertility after cancer. Hematology Am Soc Hematol Educ Program 2005;2005:523-527 [FREE Full text] [doi: 10.1182/asheducation-2005.1.523] [Medline: 16304430]

XSL•FO
RenderX

4. Reeve BB, Potosky AL, Smith AW, Han PK, Hays RD, Davis WW, et al. Impact of cancer on health-related quality of life of older Americans. J Natl Cancer Inst 2009 Jun 16;101(12):860-868 [FREE Full text] [doi: 10.1093/jnci/djp123] [Medline: 19509357]
5. McGrath P. Accommodation for patients and carers during relocation for treatment for leukaemia: a descriptive profile. Support Care Cancer 1999 Jan;7(1):6-10. [Medline: 9926967]
6. Jonker-Pool G, Hoekstra HJ, van Imhoff GW, Sonneveld DJ, Sleijfer DT, van Driel MF, et al. Male sexuality after cancer treatment--needs for information and support: testicular cancer compared to malignant lymphoma. Patient Educ Couns 2004 Feb;52(2):143-150. [Medline: 15132518]
7. Carlson LE, Angen M, Cullum J, Goodey E, Koopmans J, Lamont L, et al. High levels of untreated distress and fatigue in cancer patients. Br J Cancer 2004 Jun 14;90(12):2297-2304 [FREE Full text] [doi: 10.1038/sj.bjc.6601887] [Medline: 15162149]
8. McGrath P. Relocation for treatment for leukaemia: a description of need. Aust Health Rev 1998;21(4):143-154. [Medline: 10537552]
9. Newnham GM, Burns WI, Snyder RD, Dowling AJ, Ranieri NF, Gray EL, et al. Information from the Internet: attitudes of Australian oncology patients. Intern Med J 2006 Nov;36(11):718-723. [doi: 10.1111/j.1445-5994.2006.01212.x] [Medline: 17040358]
10. Australian Bureau of Statistics. 2009. 8146. 0 - Household Use of Information Technology, Australia, 2008-2009 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/8146.0Main+Features12008-09?OpenDocument [accessed 2011-07-20] [WebCite Cache ID 60KjyncdT]
11. US Census Bureau. 2010 Feb. Current Population Survey: Computer and Internet Use: Appendix Table A URL: http://www.census.gov/hhes/computer/ [accessed 2011-07-20] [WebCite Cache ID 60Kja3VJ6]
12. Australian Bureau of Statistics. 2007 Nov 29. 8146.0.55.001 - Patterns of Internet Access in Australia, 2006 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/8146.0.55.001Main+Features12006?OpenDocument [accessed 2011-07-20] [WebCite Cache ID 60Kigneca]
13. Gustafson DH, Hawkins R, Pingree S, McTavish F, Arora NK, Mendenhall J, et al. Effect of computer support on younger women with breast cancer. J Gen Intern Med 2001 Jul;16(7):435-445. [Medline: 11520380]
14. Gustafson DH, Hawkins R, McTavish F, Pingree S, Chen WC, Volrathongchai K, et al. Internet-Based Interactive Support for Cancer Patients: Are Integrated Systems Better? J Commun 2008 Jun;58(2):238-257. [doi: 10.1111/j.1460-2466.2008.00383.x] [Medline: 21804645]
15. Høybye MT, Dalton SO, Deltour I, Bidstrup PE, Frederiksen K, Johansen C. Effect of Internet peer-support groups on psychosocial adjustment to cancer: a randomised study. Br J Cancer 2010 Apr 27;102(9):1348-1354 [FREE Full text] [doi: 10.1038/sj.bjc.6605646] [Medline: 20424614]
16. Owen JE, Klapow JC, Roth DL, Shuster JL, Bellis J, Meredith R, et al. Randomized pilot of a self-guided internet coping group for women with early-stage breast cancer. Ann Behav Med 2005 Aug;30(1):54-64. [doi: 10.1207/s15324796abm3001_7] [Medline: 16097906]
17. Winzelberg AJ, Classen C, Alpers GW, Roberts H, Koopman C, Adams RE, et al. Evaluation of an internet support group for women with primary breast cancer. Cancer 2003 Mar 1;97(5):1164-1173 [FREE Full text] [doi: 10.1002/cncr.11174] [Medline: 12599221]
18. Curtin J. Information and Research Search Services, Department of the Parliament Library. 2001 Aug 07. A Digital Divide in Rural and Regional Australia? URL: http://www.aph.gov.au/library/pubs/cib/2001-02/02cib01.htm [accessed 2011-07-20] [WebCite Cache ID 60KhRl7o7]
19. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. 2nd edition. Sydney: Psychology Foundation; 1995.
20. Foley E, Baillie A, Huxter M, Price M, Sinclair E. Mindfulness-based cognitive therapy for individuals whose lives have been affected by cancer: a randomized controlled trial. J Consult Clin Psychol 2010 Feb;78(1):72-79. [doi: 10.1037/a0017566] [Medline: 20099952]
21. Roy-Byrne P, Craske MG, Sullivan G, Rose RD, Edlund MJ, Lang AJ, et al. Delivery of evidence-based treatment for multiple anxiety disorders in primary care: a randomized controlled trial. JAMA 2010 May 19;303(19):1921-1928. [doi: 10.1001/jama.2010.608] [Medline: 20483968]
22. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]
23. van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]
24. Warmerdam L, van Straten A, Twisk J, Riper H, Cuijpers P. Internet-based treatment for adults with depressive symptoms: randomized controlled trial. J Med Internet Res 2008;10(4):e44 [FREE Full text] [doi: 10.2196/jmir.1094] [Medline: 19033149]

XSL•FO
RenderX

25. Hinds C, Streater A, Mood D. Functions and preferred methods of receiving information related to radiotherapy. Perceptions of patients with cancer. Cancer Nurs 1995 Oct;18(5):374-384. [Medline: 7585492]
26. Rutten LJ, Arora NK, Bakos AD, Aziz N, Rowland J. Information needs and sources of information among cancer patients: a systematic review of research (1980-2003). Patient Educ Couns 2005 Jun;57(3):250-261. [doi: 10.1016/j.pec.2004.06.006] [Medline: 15893206]
27. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav 2006 Aug;9(4):494-498. [doi: 10.1089/cpb.2006.9.494] [Medline: 16901253]

Abbreviations

ARIA+: Accessibility/Remoteness Index of Australia
CI: confidence interval

Edited by G Eysenbach; submitted 22.07.11; peer-reviewed by K Eddens-Meyer, A Hartzler; comments to author 08.08.11; revised version received 08.09.11; accepted 25.09.11; published 21.12.11.

Please cite as:
Paul CL, Carey ML, Hall AE, Lynagh MC, Sanson-Fisher RW, Henskens FA
Improving Access to Information and Support for Patients With Less Common Cancers: Hematologic Cancer Patients' Views About Web-Based Approaches
J Med Internet Res 2011;13(4):e112
URL: http://www.jmir.org/2011/4/e112/
doi:10.2196/jmir.1894
PMID:22189354

©Christine Louise Paul, Mariko Leanne Carey, Alix Edna Hall, Marita Clare Lynagh, Robert W Sanson-Fisher, Frans Alexander Henskens. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 21.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Natural Language Processing to Enable In-depth Analysis of Clinical Messages Posted to an Internet Mailing List: A Feasibility Study

Tanja Bekhuis[1,2], PhD, MS, MLIS; Marcos Kreinacke[3], Dipl-Math; Heiko Spallek[2,4], PhD, DMD, MSBA(CIS); Mei Song[2], PhD; Jean A O'Donnell[5], DMD, MSN

[1]Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[2]Center for Dental Informatics, Department of Dental Public Health, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[3]Institute for Business Taxation, Leibniz University of Hanover, Hanover, Germany

[4]Office of Faculty Development and Information Management, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[5]Office of Education and Curriculum, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Tanja Bekhuis, PhD, MS, MLIS
Department of Biomedical Informatics
School of Medicine
University of Pittsburgh
UPMC Cancer Pavilion, Suite 301-338
5150 Centre Avenue
Pittsburgh, PA, 15232
United States
Phone: 1 412 647 6705
Fax: 1 412 623 4737
Email: tcb24@pitt.edu

Abstract

Background: An Internet mailing list may be characterized as a virtual community of practice that serves as an information hub with easy access to expert advice and opportunities for social networking. We are interested in mining messages posted to a list for dental practitioners to identify clinical topics. Once we understand the topical domain, we can study dentists' real information needs and the nature of their shared expertise, and can avoid delivering useless content at the point of care in future informatics applications. However, a necessary first step involves developing procedures to identify messages that are worth studying given our resources for planned, labor-intensive research.

Objectives: The primary objective of this study was to develop a workflow for finding a manageable number of clinically relevant messages from a much larger corpus of messages posted to an Internet mailing list, and to demonstrate the potential usefulness of our procedures for investigators by retrieving a set of messages tailored to the research question of a qualitative research team.

Methods: We mined 14,576 messages posted to an Internet mailing list from April 2008 to May 2009. The list has about 450 subscribers, mostly dentists from North America interested in clinical practice. After extensive preprocessing, we used the Natural Language Toolkit to identify clinical phrases and keywords in the messages. Two academic dentists classified collocated phrases in an iterative, consensus-based process to describe the topics discussed by dental practitioners who subscribe to the list. We then consulted with qualitative researchers regarding their research question to develop a plan for targeted retrieval. We used selected phrases and keywords as search strings to identify clinically relevant messages and delivered the messages in a reusable database.

Results: About half of the subscribers (245/450, 54.4%) posted messages. Natural language processing (NLP) yielded 279,193 clinically relevant tokens or processed words (19% of all tokens). Of these, 2.02% (5634 unique tokens) represent the vocabulary for dental practitioners. Based on pointwise mutual information score and clinical relevance, 325 collocated phrases (eg, *fistula filled obturation* and *herpes zoster*) with 108 keywords (eg, *mercury*) were classified into 13 broad categories with subcategories. In the demonstration, we identified 305 relevant messages (2.1% of all messages) over 10 selected categories with instances of collocated phrases, and 299 messages (2.1%) with instances of phrases or keywords for the category *systemic disease.*

XSL•FO
RenderX

Conclusions: A workflow with a sequence of machine-based steps and human classification of NLP-discovered phrases can support researchers who need to identify relevant messages in a much larger corpus. Discovered phrases and keywords are useful search strings to aid targeted retrieval. We demonstrate the potential value of our procedures for qualitative researchers by retrieving a manageable set of messages concerning systemic and oral disease.

(J Med Internet Res 2011;13(4):e98) doi:10.2196/jmir.1799

KEYWORDS

Dentistry; dental informatics; clinical research informatics; natural language processing; information storage and retrieval; electronic mail; information-seeking behavior

Introduction

In the United States, about 70% of dentists work in relative isolation as solo practitioners or in small groups [1]. Unfortunately, independent practitioners cannot afford to subscribe to all of the information resources readily available to dental faculty, academic researchers, and clinicians in large organizations. For example, the University of Pittsburgh's Health Sciences Library System [2] also serves UPMC, a global health enterprise. Dentists affiliated with either of these organizations have access to more than 3800 books on general dentistry, endodontics, pediatrics, periodontics, restoration, and special care; 15 full-text electronic books on dentistry, including important core resources; and more than 75 dentistry journals, most of which are available electronically.

In contrast, independent practitioners typically meet their information needs by relying on colleagues, discussion lists, news outlets, and a few professional journals to which they subscribe [3]. Even though most dentists in the United States have access to the American Dental Association's library by virtue of their membership, retrieval of more than the occasional full text is expensive. For example, if a member finds information in PubMed [4] not freely available in PubMed Central or an open source journal, the fee for retrieval and delivery by the library is US $7 to US $15 per article, and US $15 for one or two books, with possible late charges [5]. Fees are higher for nonmembers.

Thus, we conclude that the full panoply of important resources is inaccessible to most dentists when questions arise regarding best practice, especially at the point of care when readily available information is needed. This fact combined with dentists' preference for first consulting peers means that online communities are potentially valuable sources of information [6-8]. Such communities could be used in the future as conduits for delivery of evidence-based information, such as updated guidelines for clinical care. As for delivery of information at the point of care, this urgent need demands informatics solutions and is the focus of a US federally funded project led by Dr. Heiko Spallek [9].

Communities of Practice

An online or e-community is sometimes characterized as a virtual community of practice (CoP) [10] because members are geographically isolated yet connected socially via the Internet. A virtual CoP can serve as an information hub with easy access to expert advice and opportunities for social networking (eg, see [11]). The rationale for considering the opinions of peers expressed online is similar in spirit to the way in which research is initiated by practitioner-investigators in practice-based research networks [1,12,13]. In both cases, the value of clinical experience is recognized.

For our purposes, we are interested in knowing which clinical topics are discussed by dentists in a CoP. To do this, we mine their asynchronous messages posted to an enduring and active online discussion list. Once we understand the topics covered in the corpus of messages, we can study dentists' real information needs and the nature of their shared expertise, and can avoid delivering useless content to the community or at the point of care in future informatics applications.

Assisting Qualitative Researchers

To plan a labor-intensive study of information needs with its in-depth content analyses of clinical topics and emergent themes, one must carefully consider available human resources. For example, we have two academic dental researchers who can devote just a few days to coding and interpreting thematic content of messages with guidance from an experienced qualitative researcher. The problem then is how to assist qualitative researchers by finding a manageable number of clinically relevant messages that are worth studying given available resources.

If we know the typical length of messages, the time it takes to code a message regarding clinical topics and themes, and the number of hours researchers can devote to the content analyses, we can estimate the sample size (n) that will ensure the feasibility of the planned content analyses. Here, the corpus consisted of thousands of messages posted to an Internet mailing list for practicing dental professionals, primarily general dentists. We assumed that two academic dentists and one qualitative researcher could manage a few hundred messages.

In general, we considered three options for drawing the sample: (1) randomly sample n messages from the corpus, (2) restrict the interval of time in which n messages occur and select all messages within that interval, and (3) use natural language processing (NLP) to identify clinical topics and, depending on the research question, retrieve n messages with useful content.

The advantage of the first two options is that they are well known and easy to implement. A major disadvantage is that the selected messages may be irrelevant to the researchers' interests, especially given the informal quality of messages posted online. In the present study, the purpose of the mailing list from which the corpus originated is to offer dentists a place to discuss their clinical concerns. However, many of the messages were off

XSL•FO
RenderX

topic. For example, dentists chatted about the big football game, the trip to Europe, the swimsuit issue of *Sports Illustrated*, Michael Jackson's death, and aging parents. Although the third option is novel and more time consuming than the first two, it is in keeping with the notion that the nature of the corpus needs to be understood *before* messages are selected. This is because inferences depend on the selected units of analysis such as blocks of text [14]. Thus, the third option ensures the feasibility and probably the quality of content analyses by identifying a manageable number of messages relevant to the research question.

In this paper, we present a workflow for identifying and retrieving a manageable subset of relevant messages from a much larger corpus. It involves a sequence of machine-based steps along with human classification of clinical phrases discovered with NLP. We also demonstrate the value of this approach for enabling study of text messages by qualitative researchers. As an example, we describe the strategy we used to retrieve messages for a study underway that involves in-depth content analyses.

A preliminary version of this paper was presented at the 2010 Annual Symposium of the American Medical Informatics Association [15].

Methods

We mined the clinical content of 14,576 electronic messages posted to a fee-based discussion list during an approximate 1-year study period from April 18, 2008 to May 28, 2009. The subscribers to this global list are dental practitioners, mostly dentists from North America interested in clinical practice.

Deidentification

Because the origin of our corpus of messages is a private Internet mailing list, we took care to preserve confidentiality even though (1) the University of Pittsburgh Institutional Review Board approved this study as being exempt (PRO08040313), (2) the owner of the list deleted identifying information from the message headers before sharing content, (3) messages are regularly delivered to about 450 subscribers and then saved in a searchable archive, and (4) anyone interested in clinical dental care can subscribe. The number of subscribers and the ease with which one can subscribe suggest that this mailing list has a public aspect. Nevertheless, we went through several rounds of deidentification for two reasons: (1) to ensure confidentiality [16] for future data sharing, and (2) to optimize NLP by stripping out irrelevant information. We also used Google to confirm that excerpts presented in this paper are not easily retrievable.

During NLP (see below), we deleted stopwords (eg, articles and prepositions) to optimize discovery of topical content. Surprisingly, deletion of stopwords may help preserve anonymity. This idea is based on knowing that forensic researchers use stylistic properties of messages, including number and distribution of function or stopwords, to identify authors of email [17]. We also deleted any remaining names and places by using lists and a gazetteer, respectively, available in the Natural Language Toolkit (NLTK) [18].

Preprocessing

Mining email is challenging because of the nature of the messages [19]. For example, email can be ill formed linguistically with spelling and grammatical errors, and style can be idiosyncratic [17]. Typically, email is particularly noisy in that much of the data are irrelevant to the research question. For these reasons, processing messages is essential before clinical topics can be discovered.

Initially, we extracted the body of each message and deleted threaded responses, which is appropriate given our interest in discovering clinical topics rather than analyzing discourse. To clean the data further, we analyzed message patterns to identify recurring sources of noise (ie, data that obscure message content and meaningful frequencies in the original texts). Consequently, we deleted forwarded and quoted messages; embedded visual data such as x-ray images and photographs; virus- or spam-free notices; Microsoft Outlook notices; advertisements and footers; and signature lines. The latter often include self-promotional text.

Natural Language Processing

We used the open source NLTK version 2.0 with Python version 2.6 (Python Software Foundation, Wolfeboro Falls, NH, USA) to analyze preprocessed text. For readers new to NLP, the textbook *Natural Language Processing with Python* is a useful resource [18]. At the NLTK website [20], one can access the textbook, as well as download the programming language Python, optional packages, and the NLTK modules for NLP and text analytics.

Note that in this section we italicize terms that may be unfamiliar to readers.

We sorted and concatenated the messages by date to enable tracking discussion of topics over time. We also converted to lower case and selected *alphabetic token*s (processed words or strings of letter characters) with length >3 characters. We deleted English *stopwords* (short function words such as "a" and "the"), as well as names and places. We explored the usefulness of the obtained *vocabulary* (set of unique tokens), as well as *bigrams* and *trigrams* (pairs and triples of contiguous processed words) by examining the 100 and 300 most frequent tokens and *n-grams* (bigrams and trigrams). However, these were deemed clinically uninteresting.

To find clinical *content-bearing tokens* (substantive words such as apolipoprotein and stenosis) and phrases, we selected tokens with length >5 and frequency >7, and then derived n-grams. The rationale for this filter is similar to one presented in the NLTK text [18] where the goal is to find words and phrases that characterize a *genre*. Here the genre is *email with a clinical focus written by dental practitioners*. We also created *collocated n-grams*. *Collocations* are contiguous tokens that occur together more often than one would expect if the tokens were probabilistically independent. We selected the top 600 collocated bigrams and trigrams (300 for each type) by computing the pointwise mutual information measure for each n-gram and then sorting.

XSL•FO
RenderX

We informally confirmed that collocations derived from the content-bearing tokens were likely to retrieve useful messages by constructing *concordances* for selected tokens. A *concordance* is a set of retrieved lines with windows of text around a token or target word. The windows allow one to explore the contexts in which a target word occurs in the corpus. To build a concordance using the NLTK [20], one specifies the window size or number of characters per line, as well as the number of lines to display. For example, we examined the concordance for *lesion* to preview message content. Here are two samples from its concordance:

> *...[t]his is almost always seen in younger patients. I'm betting this lesion is of endodontic origin. Tough case to diagnose with certainty...*
>
> *...they're looking for cancer. They will NOT understand that if a lesion looks like cancer the Brush Test is not indicated. If you see a...*

Classification of Phrases and Selection of Keywords

Although most of the collocations seem to characterize dentists' clinical language, some are irrelevant. For example, here is a sample of collocations with irrelevant trigrams in italics: molecular bacterial antigens, *committing stating profitable*, *perspective agreement lobbyists*, methotrexate causative factor, inhibits demineralization enamel, *driving cadillac attack*, mutans streptococci presence.

Thus, two academic dentists (HS, JO) selected a subset of relevant collocated phrases, including bigrams and trigrams that could be used as search strings to retrieve messages with clinical content. Note that some n-grams overlap. By retaining overlapping n-grams, if they exist, we ensure a broader search than if we use just trigrams. (Most overlapping n-grams point to the same messages, but not always.) An example of an overlapping pair of n-grams is *prescribed amoxicillin hydrocodone* and *amoxicillin hydrocodone*.

The dentists also classified the phrases they selected by sorting them into broad categories with subcategories; this is considered an inductive approach to classification. Then they labeled the categories and subcategories. The process for both selection and classification was an iterative one involving discussion to reach consensus. The emergent classification scheme describes the clinical topics of concern to the dental practitioners who posted to the online mailing list. It likely will be useful to the qualitative researchers when they code messages for later content analyses [21].

After the phrases were classified, we identified embedded keywords (unigrams) to ensure that retrieval could be even broader, if desired. We defined a keyword as one that occurs at least twice in the full set of collocations. Each variant or closely related word counts as an occurrence. For example, *plaque* and *plaques*, as well as *atherosclerosis* and *atherosclerotic*, are variants; *cardiac* and *myocardial* are closely related. All six italicized examples can be used as search strings to find messages.

Finding Relevant Messages: A Demonstration

To demonstrate how the workflow presented in this paper can help researchers (see Figure 1), consider the following scenario. In our research center, a qualitative study investigating the information needs of dentists regarding the relationship between systemic disease and oral health is underway. Given this focus, two researchers independently selected some of the NLP-discovered phrases that we had identified and classified in this study. They reached consensus by discussion to determine the final list of phrases. Thus, they found a subset of phrases with embedded keywords in a subset of categories. We used the selected phrases and keywords as search strings to find messages relevant to their research question.

Because the content-bearing phrases were discovered in a merged file that had been considerably processed, a question arose as to what should be the maximum number of allowable characters between words in a phrase when searching cleaned messages not yet processed with NLP. In an informal assessment, we used 20 phrases across categories as search strings and found that the number of characters between any two words in a phrase ranged from 1 to 78. As a conservative estimate, we therefore chose to limit the interval to at most 100 characters. The aptness of this choice was borne out by the results (see below). Briefly, we carried out the following steps to retrieve and organize messages:

1. Create search strings based on collocations by first splitting phrases into words. Then for each phrase, recombine the words in any order with at most 100 characters between words. (We ignored order because words in discovered phrases were sometimes reordered in the messages, eg, *mutans streptococci* versus *streptococci mutans*.)
2. Use each keyword as a search string. If a keyword appears adjacent to another keyword in a phrase, preserve the order and search for the concatenated string.
3. Match the search strings to cleaned message texts; retrieve messages with at least one matching string.
4. Sort messages into folders (directories) per category, as well as into folders by type of match (phrase or keyword). (For example, messages with at least one phrase from the category *systemic disease* were sorted into a folder for that category, as well as a folder for all messages with instances of clinically relevant phrases. Similarly, messages with at least one keyword match were sorted into corresponding folders.)
5. Deliver deduplicated messages in folders to the researchers. (This sorting helps them find the messages they want to analyze. Further, filenames include the date when the message was posted plus a unique database identifier, which allows tracking of change in topical discussion over time, as well as retrieval of particular messages.)

For illustration purposes, consider the excerpted messages below that can be retrieved by using the following as search strings: (1) *fistula filled obturation* [trigram], (2) *herpes zoster* [bigram], and (3) *mercury* [keyword]. Remember that a maximum of 100 characters is allowed between the italicized words:

1. *...If you have a tooth with an actively draining* fistula *(pus* filled *canal), do you do one visit endo if you can get a dry*

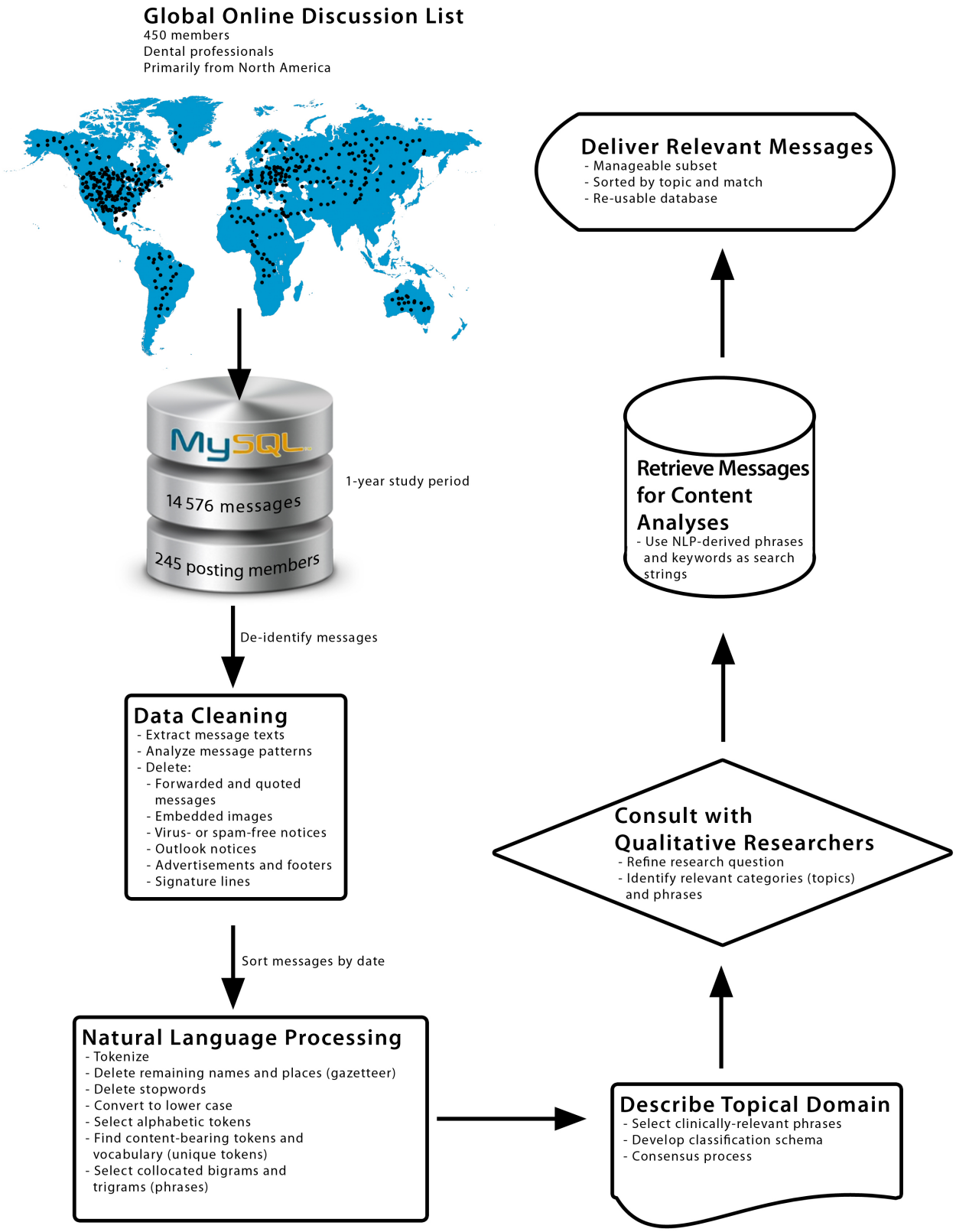
XSL•FO
RenderX

canal before obturation*? Or do you medicate for some time period and fill at a later date?...*

2. *...patient [with] recurrent ulcers on his palate [that] follow the distribution of the greater palatine nerve... I suspect* herpes zoster. *Most of the time I've seen this it's been unilateral, but in his case it's always bilateral. What other Dxs [diagnoses] should I be considering...*
3. *…Am I missing the point or is the issue (the real issue) with* mercury *not whether it causes systemic disease but rather the environmental issue of* mercury *in the food chain? We all (in the UK) have to have amalgam separators now but we know they're not foolproof...*

XSL•FO
RenderX

Figure 1. Workflow for finding clinically relevant messages posted to an Internet mailing list.



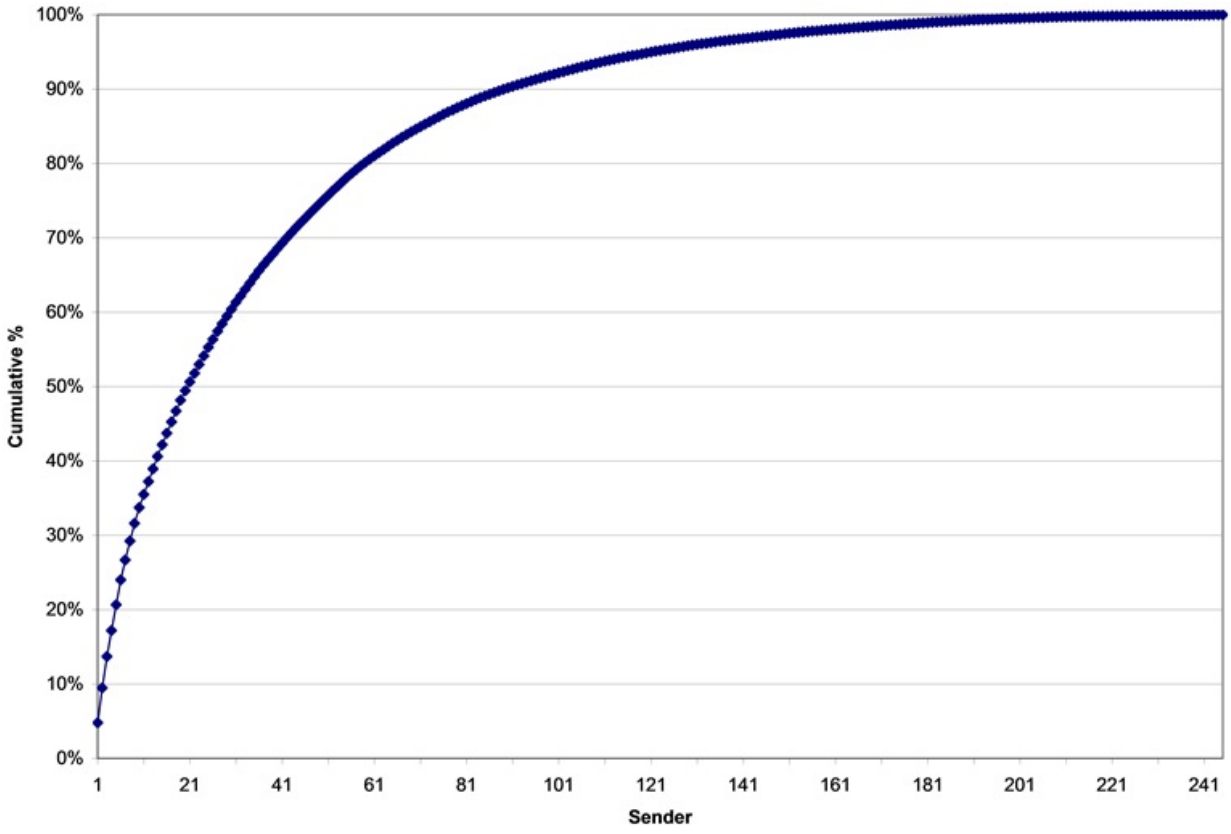
XSL•FO
RenderX

Results

Subscriber Participation

Just over half of the subscribers (245, or 54.4%) of the approximate total number of subscribers (N = 450) posted 14,576 messages. Of these, 21 subscribers (5% of the list) posted 7288 (50%) of the messages; 29 subscribers (6% of the list) posted 3644 (25%) of the messages; and 195 subscribers (43.3% of the list) posted the remaining 3644 (25%) of the messages (see Figure 2). Thus, 205 subscribers (45.6%) were passive (ie, they received messages but did not otherwise contribute to the message traffic during the study interval). Note that the total number of subscribers is approximate because the list size varies somewhat over time.

Figure 2. Cumulative distribution of messages posted by dental practitioners to an online discussion list.



Natural Language Processing

The concatenated file of cleaned messages yielded 1,468,244 tokens. Initial NLP (selecting alphabetic tokens with length >3, deleting names and places, etc) reduced the number of tokens to 533,251 (36.32%).

Filtering to find clinical content-bearing tokens yielded 279,193 tokens (19.02%). For our purposes, the unique tokens in the content-bearing set (5634, or 2.02% of the content-bearing tokens) represent the dental practitioners' vocabulary. We obtained 208,026 bigrams and 252,931 trigrams, and derived collocations. For illustration purposes, we present a handful of collocated bigrams and trigrams: *osteoclastic activity*, *painful sequestrum*, and *intravenous bisphosphonates* (bigrams); *glucose homeostasis inflammation*, *irreversible pulpitis apical*, and *supragingival scaling prophylaxis* (trigrams).

Classification of Phrases and Selection of Keywords

The classification of phrases resulted in 13 broad categories with subcategories. Table 1 presents the categories and distribution of collocated phrases and embedded keywords. The entire classification including categories and subcategories, 325 collocated phrases, and 108 embedded keywords is presented in Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Distribution of collocated phrases and keywords by category

Category	n of collocated phrases[a] (% of phrases)[b]	n of keywords[c] (% of keywords)[d]
Systemic disease	49 (15)	21 (15)
Endodontics	18 (6)	9 (6)
Orthodontics	8 (3)	3 (2)
Periodontics	12 (4)	6 (4)
Restorative dentistry	66 (20)	20 (14)
Oral and maxillofacial surgery	26 (8)	18 (13)
Other oral diseases	7 (2)	4 (3)
Radiology	7 (2)	4 (3)
Causative agent	20 (6)	9 (6)
Medication	36 (11)	19 (13)
Materials	44 (14)	17 (12)
Basic sciences	13 (4)	6 (4)
Research	19 (6)	7 (5)
Total	325	143

[a] Collocated phrases are bigrams and trigrams; selection based on pointwise mutual information score and clinical relevance.

[b] Percentage of phrases computed relative to the total number of phrases and rounded.

[c] Some keywords occur in more than one category. Thus, the total number of instances is greater than the number of unique keywords.

[d] Percentage of keywords computed relative to the total number of instances of keywords and rounded.

Finding Relevant Messages: A Demonstration

Two academic dentists conducting a qualitative study selected a subset of phrases (n_p = 144) with embedded keywords (n_{kw} = 95) in 10 of 13 categories potentially related to their research question.

Over k selected categories (k = 1 ... 10) and after deduplication, we retrieved 305 messages (range n_k, 1–119 messages) with 520 instances of matching phrases; 948 messages (range n_k, 12–343) with 1411 instances of matching keywords; and 996 messages (range n_k, 12–363) with 1931 instances of matching phrases or keywords (see Table 2). The number of characters between words in a phrase ranged from 0 to 75, after deleting white spaces and punctuation.

Table 2. Number of messages with phrases or keywords retrieved for content analyses by selected category

Selected category[a]	n of messages[b] (n of phrases)[c]	n of messages (n of keywords)	n of messages (n of phrases or keywords)
Systemic disease	119 (164)	284 (384)	299 (548)
Periodontics	14 (27)	51 (51)	54 (78)
Oral and maxillofacial surgery	36 (40)	106 (113)	106 (153)
Other oral diseases	17 (24)	44 (56)	48 (80)
Radiology	1 (1)	12 (12)	12 (13)
Causative agent	55 (78)	79 (95)	102 (173)
Medication	70 (110)	343 (377)	363 (487)
Materials	4 (4)	44 (50)	44 (54)
Basic sciences	8 (12)	157 (164)	160 (176)
Research	40 (60)	89 (109)	100 (169)
Total	305 (520)	948 (1411)	996 (1931)

[a] Categories selected from the full set by qualitative researchers.

[b] Number of messages after deduplication.

[c] Collocated phrases are bigrams and trigrams; selection based on pointwise mutual information score and clinical relevance.

XSL•FO
RenderX

To interpret Table 2, consider the row for the category *medication*. In this category, we retrieved 70 messages with 110 matches for collocated phrases, such as *intravenous bisphosphonates* from the subcategory *cancer drugs* (see Multimedia Appendix 1). We also retrieved 343 messages with 377 matches for keywords, such as *proinflammatory* from the subcategory *immune system*. Finally, we retrieved 363 messages with 487 matches for phrases or keywords selected by the dentists in the category *medication*.

Discussion

Summary of Main Findings

A workflow with a sequence of machine-based steps and human classification of NLP-discovered phrases can support researchers who need to identify relevant messages in a much larger corpus. NLP-discovered phrases and keywords are useful as search strings to aid targeted retrieval. We demonstrate the feasibility of our procedures for qualitative researchers by retrieving a manageable set of messages concerning systemic and oral disease.

Surveys Versus Textual Analysis

The reader might wonder, "Why bother with developing this workflow to support qualitative researchers? Why not survey the members of the virtual CoP and ask them outright about their information needs?"

In the research literature, studies of information needs and barriers typically focus on clinicians and primary or ambulatory care settings. Of these, just a few studies consider dentists [3,8,22,23]. So far, most of what we know is derived from survey questionnaires with items in a forced-choice format. The use of other methods is less common (eg, see [24]), even though relevant methods exist in commerce and public health. For example, marketing analysts of social media use text analytics to understand customer sentiment in unstructured text (see [25] for an accessible introduction), and researchers in infodemiology are developing mixed methods for monitoring content posted to the Internet [26,27].

Aside from the cost of developing sound surveys with appropriate sampling plans, a serious limitation is that respondents may not accurately remember the nature of their needs for evidence-based clinical information or the contexts in which needs arise. Interesting alternatives to surveys include analysis of cultural artifacts (eg, texts, images, or videos), face-to-face interviews, and field observation [28].

The investigators on our team whose project we used to demonstrate the feasibility of our procedures elected textual analysis as a way to understand clinical messages. For them, the corpus of messages posted by practicing dentists regarding specific patients or conditions is a rich data source. Appealing aspects of the corpus include the following: (1) information needs are contextually embedded, (2) messages are written in the "natural language" of dentists, and (3) discoverable clinical topics may not be what we would find with a questionnaire.

Another reason for our team's interest in textual analysis is that findings from a qualitative study can be compared with those from our own surveys (eg, see [23]), as well as from studies conducted by other teams. This will allow future assessment of threats to validity associated with method, and whether information derived from different sources is complementary.

Subscriber Participation

The very skewed distribution of subscriber participation in this study is quite similar to findings reported by Falkman et al [10], as well as Nonnecke and Preece [29]. Using the language of Wenger et al [30], Falkman and colleagues describe three groups according to their level of participation: a core group of leaders, an active group who regularly participate, and a disproportionately large group of members on the periphery. Presumably, the 5% of dental practitioners in this study who posted about half of the messages to the online discussion list were the leaders of their virtual CoP. The middle group varied considerably in their degree of participation, but they did contribute to the message traffic. Arguably, the 46% of the subscribing practitioners who never posted messages during the study interval were the peripheral group of "lurkers" or bystanders.

Interestingly, lurking on the periphery does not imply that the online community has little to offer this group. Even though passive, lurkers can still learn from core and active members who serve as information providers [10,11,30]. In fact, peripheral participation may be essential for the viability of a CoP [31] because lurking, even with its negative connotations, is "a form of participation that is both acceptable and beneficial to online groups" (p. 6, [29]).

The qualitative researchers in our group believe that clinical topics initiated and discussed by leaders and active members are probably of interest to members on the periphery. For one, they assume passive members read at least some of the messages delivered to them. They further assume that disaffected members will unsubscribe. To the extent that they are wrong, the topical domain that we have discovered may reflect the interests of core and active members rather than the entire CoP. Nevertheless, it seems reasonable to study this online dental community, as the pattern of participation is typical of other communities of practice and electronic discussion lists.

Natural Language Processing

To cope with the noisy and informal nature of email, we heavily processed the messages. In so doing, we may have inadvertently overlooked important content-bearing phrases by deriving collocations from a much-reduced set of tokens. Nevertheless, collocations are much more informative than frequent phrases [18]. The latter are usually uninteresting, at least in this context, and seem to derive from ordinary language, repeated self-promotion, and banner advertisements. Despite our best efforts, we were unable to delete all of the text-based noise.

Many of the messages include excerpts from news items, magazine articles, or research articles. These excerpts seem to have a disproportionate number of clinical phrases relative to message content written by subscribers. (Chew and Eysenbach [26] identified a similar problem when analyzing the content of posts to Twitter ("tweets"; see [32]) during the 2009 H1N1 pandemic. They cautioned that key phrases in spam and popular

XSL•FO
RenderX

news might affect retrieval of tweets and activity over time.) Because we were unable to identify automatically all of the imported content, we analyzed the entire message after preprocessing. However, one could argue that members, especially leaders, bring in relevant text and that mining messages with imported text still leads to a reasonable set of NLP-derived phrases.

Finding Relevant Messages: A Demonstration

In this study, we demonstrate the potential usefulness of our procedures by retrieving a manageable set of relevant messages for qualitative researchers. Their research entails exploring dentists' knowledge of the relationship between systemic and oral disease expressed in messages. To understand how they can work with messages sorted by category and type of match, consider the following scenario.

Assume the researchers can handle about 300 messages for labor-intensive content analyses. They could design a broad or focused study by considering the number and type of match in each category. For example, for a broad study, they could analyze the 305 messages with clinical content-bearing phrases that we retrieved for the categories they had selected (see Table 2). For a more focused study, they could elect to work with messages from just the first category, *systemic disease*, which has 299 messages with 548 instances of phrases or keywords. Alternatively, they could select messages in some other combination of categories and type of match with the constraint that the total number of messages to analyze is about 300. If they decide to add a clinician to the team or devote more time to the project, they could analyze a larger set of messages.

By sorting the messages we retrieved into the categories selected a priori by the qualitative researchers, we were able to create a useful database that encourages flexible investigation.

Limitations

A major limitation of this study is that we used a single source to mine electronic messages. It is possible that the NLP-discovered phrases and their subsequent classification will not generalize to other communities. In other words, the topical domain that we discovered may not describe the clinical interests of other practitioners, such as dentists who prefer to remain offline. Even if our version of the topical domain is useful, we still need to assess whether and how it changes over time. Additionally, other methods such as latent semantic analysis, sometimes referred to as latent semantic indexing [33-35], could yield a different set of topics. Finally, although we took care to reach consensus when classifying phrases, other dental researchers could have seen a different structure. Nevertheless, the limitations of any feasibility study are offset by the potential for usefulness and discovery. We believe the limitations of this study can be addressed in the future with formal evaluations that compare methods and communities.

Future Research

Each step in the workflow presents opportunities for further research. Nevertheless, once the system we are developing becomes reasonably efficient and robust, a cost-benefit analysis will be appropriate. For example, we could compare the labor involved and quality of retrieval for a simple random sample of messages with ad hoc keyword searches as a baseline versus our system.

Other methods to identify clinically relevant messages, such as summarization and clustering of similar summaries [19,36,37], or use of an ontology to enable retrieval (eg, see [38]) could be worthwhile. Also, discourse analysis [18] of the threaded messages could help us better understand how clinicians respond to the information needs of their peers, and whether the shared information is in keeping with the best evidence in published guidelines.

Ultimately, this program of research will help us improve knowledge transfer of useful information for the legions of dentists who practice in relative isolation.

Acknowledgments

We would like to thank David S. Dodell, DMD, for his help in obtaining the messages posted to the discussion list. The Pittsburgh Biomedical Informatics Training Program 5T15LM007059 partially supported this research.

Conflicts of Interest

None declared

Multimedia Appendix 1

Classification of dental phrases with keywords. A table of keywords by category is displayed at the end of the classification.

[PDF file (Adobe PDF File), 149 KB - jmir_v13i4e98_app1.pdf]

References

1. Curro FA, Craig RG, Van Thompson P. Practice-based research networks and their impact on dentistry: creating a pathway for change in the profession. Compend Contin Educ Dent 2009 May;30(4):184, 186-184, 187. [Medline: 19441734]
2. anonymous. Health Sciences Library System. 2011. Pitt Resources Quick Search URL: http://www.hsls.pitt.edu/ [accessed 2011-05-11] [WebCite Cache ID 5ybUzxPzl]

XSL•FO
RenderX

3. Song M, Spallek H, Polk D, Schleyer T, Wali T. How information systems should support the information needs of general dentists in clinical settings: suggestions from a qualitative study. BMC Med Inform Decis Mak 2010;10:7 [FREE Full text] [doi: 10.1186/1472-6947-10-7] [Medline: 20122272]
4. PubMed.gov. US National Library of Medicine, National Institutes of Health. 2011. pubMed URL: http://www.ncbi.nlm.nih.gov/pubmed/ [accessed 2011-05-11] [WebCite Cache ID 5ybVxMtSC]
5. American Dental Association. ADA. 2011. Fees for Members URL: http://www.ada.org/3791.aspx [accessed 2011-05-11] [WebCite Cache ID 5ybXRHA5P]
6. Landry CF. Work roles, tasks, and the information behavior of dentists. J Am Soc Inf Sci Technol 2006;57(14):1896-1908. [doi: 10.002/asi.20385]
7. Schleyer TK, Forrest JL, Kenney R, Dodell DS, Dovgy NA. Is the Internet useful for clinical practice? J Am Dent Assoc 1999 Oct;130(10):1501-1511 [FREE Full text] [Medline: 10570599]
8. Strother EA, Lancaster DM, Gardiner J. Information needs of practicing dentists. Bull Med Libr Assoc 1986 Jul;74(3):227-230. [Medline: 3742116]
9. Spallek H. National Institue of Dental & Craniofacial Research, National Institutes of Health, Research Portfolio Online Reporting Tools (RePORT). 2011 May 11. Implementing Research Findings and Evidence-Based Interventiona Into Real-World Dental Practice Settings URL: http://projectreporter.nih.gov/project_info_description.cfm?aid=8028402&icde=8077609 [WebCite Cache ID 5ybZrgg6B]
10. Falkman G, Gustafsson M, Jontell M, Torgersson O. SOMWeb: a semantic web-based system for supporting collaboration of distributed medical communities of practice. J Med Internet Res 2008;10(3):e25 [FREE Full text] [doi: 10.2196/jmir.1059] [Medline: 18725355]
11. Spallek H, Irwin JY, Schleyer T, Butler BS, Weiss PM. Supporting the emergence of dental informatics with an online community. Int J Comput Dent 2007 Jul;10(3):247-264. [Medline: 18271498]
12. Gilbert GH, Williams OD, Rindal DB, Pihlstrom DJ, Benjamin PL, Wallace MC, DPBRN Collaborative Group. The creation and development of the dental practice-based research network. J Am Dent Assoc 2008 Jan;139(1):74-81 [FREE Full text] [Medline: 18167389]
13. Lindbloom EJ, Ewigman BG, Hickner JM. Practice-based research networks: the laboratories of primary care research. Med Care 2004 Apr;42(4 Suppl):III45-III49. [Medline: 15026664]
14. Roberts CW. Content analysis. In: Smelser NJ, Baltes PB, editors. International Encyclopedia of the Social & Behavioral Sciences. Oxford, UK: Pergamon; 2001:2697-2702.
15. Bekhuis T, Kreinacke M, Spallek H, Song M. Using the Natural Language Toolkit to reduce the number of messages for in-depth content analyses: a case study. In: AMIA Annual Symposium Proceedings. 2010 Presented at: AMIA 2010; Nov 13-17, 2010; Washington, DC, USA.
16. 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]
17. de Vel O, Anderson A, Corney M, Mohay G. Mining e-mail content for author identification forensics. ACM SIGMOD Rec 2001;30:55-64. [doi: 10.1145/604264.604272]
18. Bird S, Klein E, Loper E. Natural Language Processing With Python. Sebastopol, CA: O'Reilly; 2009.
19. Katakis I, Tsoumakas G, Vlahavas I. Email mining: emerging techniques for email management. In: Vakali A, Pallis G, editors. Web Data Management Practices: Emerging Techniques And Technologies. Hershey, PA: Idea Group Publishing; 2006:219-240.
20. Natural Language Toolkit. Home page URL: http://www.nltk.org/ [accessed 2011-05-09] [WebCite Cache ID 5yYGW5pg6]
21. Dilevko J, Gottlieb L. The relevance of classification theory to textual analysis. Libr Inf Sci Res 2009;31:92-100. [doi: 10.1016/j.lisr.2009.01.001]
22. Duxbury AJ, Leach FN. Drug information and the dental practitioner. Dent Update 1981 Mar;8(2):101-2, 106. [Medline: 6943113]
23. Spallek H, Song M, Polk DE, Bekhuis T, Frantsve-Hawley J, Aravamudhan K. Barriers to implementing evidence-based clinical guidelines: a survey of early adopters. J Evid Based Dent Pract 2010 Dec;10(4):195-206. [doi: 10.1016/j.jebdp.2010.05.013] [Medline: 21093800]
24. Freimuth VS, Massett HA, Meltzer W. A descriptive analysis of 10 years of research published in the Journal of Health Communication. J Health Commun 2006;11(1):11-20. [doi: 10.1080/10810730500461042] [Medline: 16546916]
25. Sterne J. SAS Institute. 2010. Text Analytics for Social Media: Evolving Tools for an Evolving Environment URL: http://www.sas.com/resources/whitepaper/wp_24091.pdf [accessed 2011-09-30] [WebCite Cache ID 625lHnY2R]
26. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
27. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
28. Green J, Thorogood N. In: Green JA, Thorogood N, editors. Qualitative Methods for Health Research (Introducing Qualitative Methods series). 2nd edition. Los Angeles, CA: Sage; 2009.

XSL•FO
RenderX

29. Nonnecke B, Preece J. Lurker demographics: counting the silent. In: Proceedings of CHI 2000: The Future is Here.: Association of Computing Machinery (ACM); 2000 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 1-6, 2000; Hague, Netherlands. [doi: 10.1145/332040.409]
30. Wenger E, McDermott R, Snyder WM. Cultivating Communities of Practice: A Guide to Managing Knowledge. Boston, MA: Harvard Business School Press; 2002.
31. Spallek H, Butler BS, Schleyer TK, Weiss PM, Wang X, Thyvalikakath TP, et al. Supporting emerging disciplines with e-communities: needs and benefits. J Med Internet Res 2008;10(2):e19 [FREE Full text] [doi: 10.2196/jmir.971] [Medline: 18653443]
32. Twitter, Inc. 2011. twitter URL: http://twitter.com/ [accessed 2011-05-22] [WebCite Cache ID 5ysGmdFEz]
33. Bellegarda J. Exploiting latent semantic information in statistical language modeling. Proc IEEE 2000;88(8):1279-1296. [doi: 10.1109/5.880084]
34. Hand D, Mannila H, Smyth P. Principles of Data Mining. Cambridge, MA: MIT Press; 2001.
35. Manning CD, Raghavan P, Schutze H. Introduction to Information Retrieval. New York, NY: Cambridge University Press; 2008.
36. Radev DR, Teufel S, Saggion H, Lam W, Blitzer J, Qi H, et al. Evaluation challenges in large-scale document summarization. In: ACL '03. 2003 Presented at: 41st Annual Meeting of the Association for Computational Linguistics; July 7-12, 2003; Sapporo, Japan. [doi: 10.3115/1075096.1075144]
37. Saggion H, Gaizauskas R. Multi-document summarization by cluster/profile relevance and redundancy removal. In: Proceedings. 2004 Presented at: Document Understanding Conference (DUC04) 2004; May 6-7, 2004; Boston, MA, USA.
38. Konovalov S, Scotch M, Post L, Brandt C. Biomedical informatics techniques for processing and analyzing web blogs of military service members. J Med Internet Res 2010;12(4):e45 [FREE Full text] [doi: 10.2196/jmir.1538] [Medline: 20923755]

Abbreviations

CoP: community of practice
NLP: natural language processing
NLTK: Natural Language Toolkit

Edited by G Eysenbach; submitted 10.03.11; peer-reviewed by A Archaya, Q Zeng, A Keselman; comments to author 19.04.11; revised version received 24.05.11; accepted 08.07.11; published 23.11.11.

Please cite as:
Bekhuis T, Kreinacke M, Spallek H, Song M, O'Donnell JA
Using Natural Language Processing to Enable In-depth Analysis of Clinical Messages Posted to an Internet Mailing List: A Feasibility Study
J Med Internet Res 2011;13(4):e98
URL: http://www.jmir.org/2011/4/e98/
doi:10.2196/jmir.1799
PMID:22112583

©Tanja Bekhuis, Marcos Kreinacke, Heiko Spallek, Mei Song, Jean A O'Donnell. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

WikiBuild: A New Online Collaboration Process For Multistakeholder Tool Development and Consensus Building

Samir Gupta[1,2,3], MSc, MD; Flora T Wan[2], MASc; David Newton[2], MA; Onil K Bhattacharyya[2,4], MD, PhD; Mark H Chignell[5], PhD; Sharon E Straus[1,2,6], MSc, MD

[1]Department of Medicine, University of Toronto, Toronto, Canada

[2]The Keenan Research Centre, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada

[3]Division of Respirology, Department of Medicine, St. Michael's Hospital, Toronto, ON, Canada

[4]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[5]Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada

[6]Division of Geriatrics, Department of Medicine, St Michael's Hospital, Toronto, ON, Canada

Corresponding Author:
Samir Gupta, MSc, MD
Division of Respirology
Department of Medicine
St. Michael's Hospital
Suite 6045, Bond Wing
30 Bond St
Toronto, ON, M5B 1W8
Canada
Phone: 1 416 864 6026
Fax: 1 416 864 6026
Email: guptas@smh.ca

Related Article:

Comment in: Archambault PM. WikiBuild: A New Application to Support Patient and Health Care Professional Involvement in the Development of Patient Support Tools. J Med Internet Res. 2011;13(4) p. e114 http://www.jmir.org/2011/4/e114/

Abstract

Background: Production of media such as patient education tools requires methods that can integrate multiple stakeholder perspectives. Existing consensus techniques are poorly suited to design of visual media, can be expensive and logistically demanding, and are subject to caveats arising from group dynamics such as participant hierarchies.

Objective: Our objective was to develop a method that enables multistakeholder tool building while averting these difficulties.

Methods: We developed a wiki-inspired method and tested this through the collaborative design of an asthma action plan (AAP). In the development stage, we developed the Web-based tool by (1) establishing AAP content and format options, (2) building a Web-based application capable of representing each content and format permutation, (3) testing this tool among stakeholders, and (4) revising this tool based on stakeholder feedback. In the wiki stage, groups of participants used the revised tool in three separate 1-week "wiki" periods during which each group collaboratively authored an AAP by making multiple online selections.

Results: In the development stage, we recruited 16 participants (9/16 male) (4 pulmonologists, 4 primary care physicians, 3 certified asthma educators, and 5 patients) for system testing. The mean System Usability Scale (SUS) score for the tool used in testing was 72.2 (SD 10.2). In the wiki stage, we recruited 41 participants (15/41 male) (9 pulmonologists, 6 primary care physicians, 5 certified asthma educators, and 21 patients) from diverse locations. The mean SUS score for the revised tool was 75.9 (SD 19.6). Users made 872, 466, and 599 successful changes to the AAP in weeks 1, 2, and 3, respectively. The site was used actively for a mean of 32.0 hours per week, of which 3.1 hours per week (9.7%) constituted synchronous multiuser use (2–4 users at the same time). Participants averaged 23 (SD 33) minutes of login time and made 7.7 (SD 15) changes to the AAP per day. Among participants, 28/35 (80%) were satisfied with the final AAP, and only 3/34 (9%) perceived interstakeholder group hierarchies.

XSL•FO
RenderX

Conclusion: Use of a wiki-inspired method allowed for effective collaborative design of content and format aspects of an AAP while minimizing logistical requirements, maximizing geographical representation, and mitigating hierarchical group dynamics. Our method faced unique software and hardware challenges, and raises certain questions regarding its effect on group functioning. Potential uses of our method are broad, and further studies are required.

(J Med Internet Res 2011;13(4):e108) doi:10.2196/jmir.1833

KEYWORDS

Consensus; focus groups; user-computer interface; Web 2.0; asthma; self-care

Introduction

Objective

We sought to develop and test an innovative wiki-inspired technology to facilitate collaborative design and consensus building across multiple stakeholders. In particular, our method aims to enable multiuser development of the content and format of media such as patient education tools. We developed and tested this technology through the collaborative design of an asthma self-management tool called an asthma action plan (AAP).

Background And Significance

An AAP is an individualized written plan produced by a health care professional (HCP) for a patient with asthma, for the purpose of providing education and guidelines for self-management of worsening asthma symptoms [1]. Because most AAP templates have been developed from a predominantly expert medical rather than from a primary care physician (PCP) or patient perspective [2], both AAP delivery by PCPs and AAP use by patients remains low [2,3], despite strong evidence for their benefit [1].

Development methods that integrate the perspectives of all relevant stakeholders have been shown to be more likely to yield products that are appropriate to the local context and that effectively meet the needs of end users [4]. Accordingly, stakeholder engagement in the development process is a key determinant of the implementability of products such as evidence summaries [5,6], guidelines [7], and patient tools [8]. Accordingly, we sought to develop a more readily implementable AAP through a design process that would enable inclusion of multiple stakeholders, including PCPs and patients.

The three main formal consensus techniques used in health care are the Delphi method, the nominal group technique (NGT), and the consensus development conference [9]. The Delphi method consists of questionnaires mailed serially to stakeholders, interim feedback mailed to individuals regarding group preferences, and aggregation of responses according to explicit statistical principles [9,10]. The “online Delphi” technique applies the same principles; however, participants complete questionnaires electronically and can be linked through a central computer that continually updates and displays group preferences to individual participants [10,11]. In the NGT, participants are assembled face-to-face, each records his or her ideas independently, and these ideas are listed for the group and discussed with the help of a facilitator. Individual judgments and votes are recorded and aggregated statistically to derive the group judgment. Finally, a consensus development conference consists of a moderated, unstructured meeting where evidence and ideas are presented by various interest groups or experts who are not members of the decision-making group, and the latter retire to attempt to reach consensus. Both the open and the private group discussions are chaired, and implicit methods are used to aggregate opinions [9].

These techniques present several disadvantages for development of an AAP. First, an AAP is a visual medium. Previous studies have demonstrated the advantage of incorporating human factors design elements in visual media intended for health care interventions [12]. However, existing consensus and focus group techniques are poorly suited to achieving agreement about the physical attributes (format) of visual media [9] or for novel visual media design, due to inherent difficulties in expressing aesthetic preferences and describing imagined visual characteristics verbally [4]. Second, in-person techniques have a number of practical limitations. A skilled moderator is required [13,14], and may be difficult and expensive to access [6]. Other costs include travel and accommodation for stakeholders. Planning requires organizational support, and recruitment of appropriate participants can be challenging due to conflicting schedules [5,6]. Finally, in-person techniques are subject to complex group dynamics. Participants may be pressured to agree with a group’s or a dominant individual's viewpoint [9,15], and social hierarchies may form, favoring professionals over patients. Some individuals may not articulate their preferences due to a lack of confidence, a lack of trust in the group, or poor group management by the moderator [9,15].

We conceived of a novel wiki-inspired method to achieve both consensus and collaborative design. A wiki is a hypertext-based collaborative software that allows users not only to add content, but also to edit and alter existing content according to their preferences. Wikis have been used for collaborative writing, but not for development of visual media [16]. In medical research, wikis have been used to support the implementation of an electronic medical record system [17] and to build online catalogues of genetic codes, protein structure [18,19], medical ontology [20], and medical information [21,22]. Use of a wiki platform to ascertain and summarize the preferences of multiple users or in the design of a medical communication tool such as an AAP has not previously been reported.

Methods

We developed and tested a system that allows multiple users to collaboratively design an AAP by inputting preferences for the content and format (visual layout and design) of the AAP through a Web-based wiki-inspired platform. In order to

XSL•FO
RenderX

accurately reflect format permutations, users constructed and viewed the AAP in real time, as they navigated a series of choices in drop-down menus. All elements of the study were approved by our institutional review board.

Development Stage

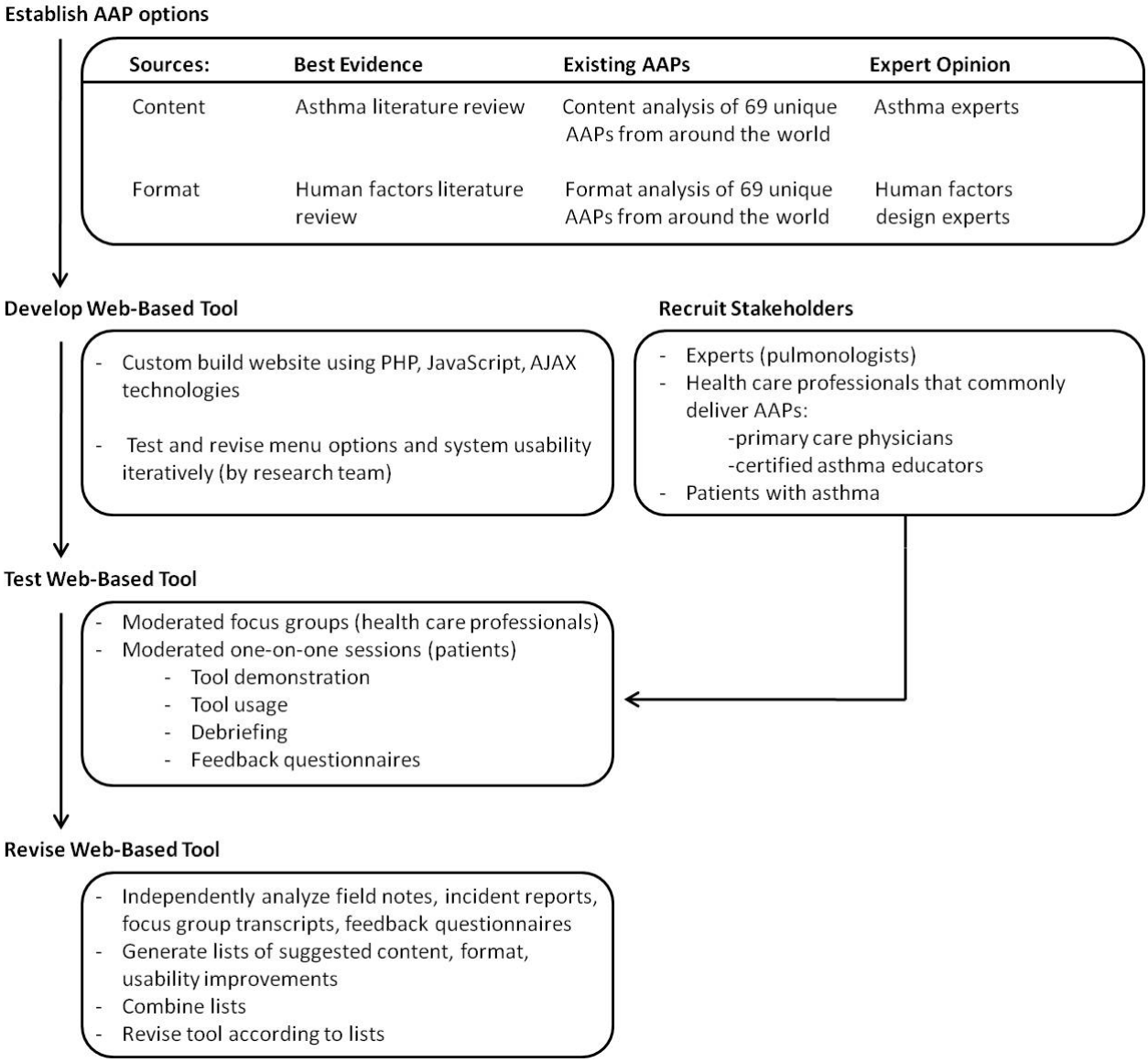
Development occurred in 4 steps (Figure 1). First, we established content and format options to include in the system, using best evidence from medical and human factors literature, a review of 69 existing AAPs collected from around the world, and opinions from asthma and human factors experts. Second, we built a Web-based application to enable representation of each content and format permutation (for the Safari Web browser, version 5.0.4; Apple Inc, Cupertino, CA, USA). Members of the research team serially tested and revised the system. Third, the tool was tested by stakeholders to optimize content and format choices and system usability. Relevant stakeholders were asthma experts (pulmonologists), the HCPs who commonly deliver AAPs (PCPs and certified asthma educators [CAEs]), and patients with asthma. Participants were purposively sampled to reflect hospital and community practice settings. Pulmonologists, CAEs, patients, and PCPs were recruited from a quaternary care hospital (St. Michael's Hospital, Toronto, Ontario, Canada), and PCPs and patients were recruited from community clinics within the Greater Toronto Area. Patient inclusion criteria included a self-reported physician diagnosis of asthma, ability to speak English, and adequate computer skills (as determined by a brief screening questionnaire; please see the supplementary table in Multimedia Appendix 1).

We conducted 2-hour focus group sessions for HCPs (2–4 participants each) and 2-hour individual sessions for patients. Each session was facilitated by a moderator and attended by the study coordinator and computer programmer, who undertook troubleshooting, and by 2 study investigators, who took field notes. Sessions were audio recorded and the tapes transcribed verbatim. Each session was scripted and began with a presentation providing background on AAPs and the purpose of the study, followed by a tool demonstration and a 45-minute period during which participants were asked to individually use the tool to develop a "best possible" (blank) AAP. Study investigators observed each participant and documented difficulties on a standardized case report form. Copies of each participant's final AAP were printed and distributed to all participants. In a group debriefing session, we discussed each case report to elucidate problems and corresponding improvements, and sought feedback on system usability and choices in each menu. Participants completed an online questionnaire consisting of a series of statements with 5-point Likert scales measuring agreement, open-ended questions, and the System Usability Scale (SUS) [23]. Participants were reimbursed for their time.

Finally, 2 members of the research team independently analyzed all field notes, case reports, focus group transcripts, and online feedback. Each member generated a list of suggested changes to content and format options and usability features of the tool. We revised the tool based on these suggestions.

XSL•FO
RenderX

Figure 1. Method development process used in developing our wiki tool: (1) establish content and format options to include in the system, (2) build a Web-based application to enable representation of each content and format permutation through iterative revisions by the research team, while concurrently recruiting participants for the next stage, (3) test the tool in each stakeholder group to optimize content and format choices and system usability, (4) revise the tool based on data from the testing stage.



Wiki Stage

We used this revised tool for collaborative design in a wiki environment (tool available online at http://knowledgetranslation.ca/octapus_i/login.php?access=guest). This tool was inspired by the wiki concept and was similar to conventional wikis in the following ways: Web-based; used collaboratively by multiple users; invited all users to add edits; did not require any browser add-ons for core site functions; acted as a database for creating, browsing, and searching information; allowed for nonlinear, evolving, complex, and networked text, argument, and interaction; enabled real-time webpage creation and updating (without review before modifications were accepted and displayed online); and enabled a natural selection process to guide site content [24]. However, our technology also differed from conventional wikis in the following important ways: did not make use of simplified markup language or a "wysiwyg" text editor; did not invite casual users to be part of the wiki process; constrained edits by offering predetermined options rather than "free text" editing; offered users the ability to edit visual characteristics (format) of the website itself, rather than text content exclusively; and was not powered by wiki software (we used a custom-built platform rather than the MediaWiki software). Our application was custom built on the following frameworks: jQuery, version 1.3.1 (a JavaScript library with built-in AJAX functions was used for the client-side interaction); wkpdf, version 0.2 (used for PDF generation); PHP, version 5.2 (including PEAR and MDB2) (used for server-side functionality); and MySQL, version 5.1 (used for databases; Oracle Corporation, Redwood Shores, CA, USA).

We recruited 3 groups of new users, each composed of 14 participants (3 pulmonologists, 2 PCPs, 2 CAEs, and 7 patients with asthma) sampled purposively to reflect a broad range of settings. HCPs were recruited from hospitals and community clinics in Canada, from hospitals in the United States and Australia, from a Canadian AAP workshop, and through the

XSL•FO
RenderX

Ontario Lung Association. Patients were recruited from hospitals and community clinics within the Greater Toronto Area and through the Asthma Society of Canada. Patients required a self-reported physician diagnosis of asthma, and all participants had to fulfill the requirements listed in the supplementary table in Multimedia Appendix 1 as well as the following: (1) access to high-speed Internet at work or at home, (2) average weekly Internet use at work or home of ≥4 hours, and (3) minimum once-weekly use of at least three of the following applications: email, Internet Explorer, Mozilla Firefox, Apple Safari, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, or Adobe PDF.

We conducted orientation sessions 1 week before the start of each wiki session, summarizing AAPs, the study's purpose, and the system's functions. HCPs received this orientation through a 1-hour moderated Livestream webinar (Livestream, New York NY, USA) and patients through moderated face-to-face group sessions (2–4 participants each). Each user was asked to download Safari and to confirm the tool's function on his or her computer before these sessions.

Each 14-participant group was given a 1-week period to collaboratively author a single AAP. Participants collaborated through the site's wiki function, whereby any member could alter online choices made previously by other members. Participants received daily reminders to use the tool. The tool included a log of previously made choices, a chat room for online discussions, and comment fields attached to each choice, enabling members to propose supporting arguments for their choices. In the event of an ''edit war" [16], defined as 14 serial changes to a single menu option made by 2 participants over a 48-hour period, the tool automatically triggered an online vote of all group members. The result from this vote would determine the option choice. Users had 24-hour access to the site and to technical support through email and telephone.

At the end of each 1-week period, participants received their group's final AAP and completed an online questionnaire measuring perceptions of the tool, the wiki process, and the AAP. We documented logistical and technical difficulties associated with the technology and analyzed tool usage. We used expert opinion to define the following criteria for a successful wiki process: (1) high usage rates (a mean of ≥10 minutes of active logged-in time per user per day), (2) positive measures of usability (a higher mean SUS score than in the development stage, and a mean SUS score ≥72.5, corresponding to "good" or better usability) [25], and (3) high user satisfaction with the final product (a decreasing trend for changes made through the week-long process, and ≥75% user agreement with questionnaire statements relating to satisfaction with the final AAP).

Results

Development Stage

For testing, we recruited 16 participants (9/16 male) (4 pulmonologists, 4 PCPs, 3 CAEs, and 5 patients). Of the 16 participants, 7 (44%) were between 30 and 39, 4 (25%) were between 40 and 49, 4 were (25%) between 50 and 59, and 1 (6%) was ≥60 years of age. Of the 5 patients, 1 (20%) had a high school education, 1 (20%) had a college or trade school education, and 3 (60%) had a university education. The mean SUS was 72.2 (SD 10.2): 75.0 (SD 8.16) for pulmonologists; 76.2 (SD 11.1) for PCPs; 66.7 (SD 5.77) for CAEs; and 70.0 (SD 13.5) for patients (scores for the SUS can range from 0 to 100). We made several significant usability-related changes to the system on the basis of feedback received in the focus groups and interviews in the development stage. Table 1 presents user comments and corresponding revisions.

XSL•FO
RenderX

Table 1. Usability comments and corresponding improvements to the wiki-based asthma action plan (AAP) development system

Difficulty/criticism	Change made to system
Users ran out of space to add items to the AAP without realizing it	Added a pop-up warning for users when space was exceeded
	Added a hover above each layout to indicate approximate number of options that could be added in each AAP zone, with each layout
Users felt that their ability to experiment with different 3- vs 4-zone formats and a 1- vs 2-step yellow zone was limited by the fact that information added to the extra zone/step was lost if they switched formats	Reprogrammed the system such that items last contained in the orange zone or second step of the yellow zone would reappear if the zone/step was added back after having been removed
Users did not realize when a menu had ended and when to move to the next menu	Added a statement at the end of each menu section indicating that it is complete and directing users to the next appropriate section
Users did not find the use of arrows intuitive for opening and closing menus	Arrows were changed to "+" and "–" signs used in conventional Windows navigation
Users did not realize that choosing options in certain menus would automatically open further submenus containing phrases to complete these (otherwise nonsensical) statements	The following note was added below menu titles for all menus containing any such submenus: "*(note: selecting this option may produce further submenus with more options for you to choose)*"
When trying to remove or alter an existing item in the AAP, users often found it difficult to find the corresponding menu	Added a function such that double clicking on a selection in the AAP would open up the corresponding content root menu in the menu window
Users wanted to "line up" similar items across zones and control the order of items in each zone description and instruction area	Added a function such that items in each of the zone description and instruction areas could be reordered by clicking and dragging directly in the AAP window
Several users did not scroll down to see all menu options when menus were opening downward (these were hidden by the menu window box and required scrolling)	Reprogrammed menus such that each menu could be seen rising to the top of the menu window box once expanded, displaying scrollable choices below
Users indicated that they required more screen space to view the AAP (particularly laptop users)	Added a function enabling users to temporarily hide menus and to click and drag the entire AAP to the center of the screen for viewing
Users preferred to have similar titles for each zone and found it cumbersome to choose these separately in each zone menu	Added zone title selections to the "Setup" tab, such that these choices could be made consecutively

Wiki Stage

We recruited 41 participants (15/41 male) (9 pulmonologists, 6 PCPs, 5 CAEs, and 21 patients) from 16 different cities, 5 Canadian provinces (Ontario, Quebec, Manitoba, Alberta, and British Columbia), and 3 countries (Canada, United States, and Australia). Although our target was 42 participants, 1 CAE who was recruited to the second wiki week withdrew, and we were unable to successfully recruit another participant in time for that week. This CAE later successfully participated in the third wiki week.

Among 39 participants whose age was available, 3 (8%) were <30, 10 (26%) were between 30 and 39, 14 (36%) were between 40 and 49, 9 (23%) were between 50 and 59, and 3 (8%) were ≥60 years of age. Of the 21 patients, 2 (10%) had a high school education, 10 (48%) had a college or trade school level education, and 9 (43%) had a university education. Our analysis focused on participation, system access, system usage, and user perceptions in the wiki stage.

Participation

Of the 42 target participants, 7 (17%) did not participate in the process: 3 missed the training seminar, due to a family emergency (1), sudden illness (1), and inability to access Livestream due to university firewalls (1); 2 could not download Safari due to university firewalls; 1 did not register for the site after training; and 1 did not log in despite technical assistance. Of the 42 target participants, 5 (12%) reported reasons for limited participation: 2 could not download Safari to an office computer due to firewalls but accessed the site from home; 1 had computer problems that limited participation; 1 could not download Safari to an office computer and had problems on the home computer, and 1 was hospitalized for 3 days of the 7-day wiki process. Lost and limited participants were approximately evenly distributed across wiki weeks and participant types, and all 3 wiki weeks had full-time participation from at least one of each user type.

System Usage

Of 347 login attempts over the 3 wiki weeks, 128 (36.9%) failed due to use of incorrect browsers. With the help of technical support personnel (mostly through email communication), all but 1 participant eventually successfully accessed the site through Safari. A total of 872, 466, and 599 successful changes to the AAP were made in wiki weeks 1, 2, and 3, respectively. Of these 1937 changes, 453 (23.4%) related to AAP format and 1484 (76.6%) to content, and no edit wars occurred. A video demonstrating the evolution of the AAP over the first wiki week is available in Multimedia Appendix 2. One PCP (week 2), and 1 pulmonologist (week 3) logged in but did not make changes to the AAP. The mean number of conversations (≥2 participants exchanging chat messages) was 8.0/week, with an average of 5.8 messages and 2.8 participants per conversation. Through comments and chats, 6 of 19 (32%) patients and 7 of 17 (41%) HCPs (total 13/36, 36%) revealed the stakeholder group to which they belonged. The site was used actively for a mean of

XSL•FO
RenderX

32.0 hours per week, of which 3.1 hours/week (9.7%) constituted synchronous multiuser use (2–4 users at the same time). Table 2 details the system usage data.

Table 2. System usage data

	Participant type				
	Pulmonologists	Primary care physicians	Certified asthma educators	Patients with asthma	All users
Mean (SD) number of days logged in/week/user	3.1 (2.0)	4.4 (1.5)	5.2 (1.6)	5.3 (2.0)	4.7 (2.0)
Mean (SD) number of logins/week/user	4.4 (5.8)	5.8 (5.2)	6.4 (6.2)	6.7 (6.6)	6.1 (6.3)
Mean (SD) active logged-in time/day/user[a] (minutes)	14 (24)	16 (23)	38 (50)	24 (31)	23 (33)
Mean (SD) number of changes made/day/user	6.3 (14)	6.0 (12)	7.2 (9.9)	8.8 (17)	7.7 (15)
Range of total number of changes made/user (in 1 week)	0–148	0–146	10–102	5–357	0–357
Mean (SD) number of changes made/login	9.9 (21)	7.7 (13)	8.0 (10)	9.3 (17)	9.0 (16)
Mean (SD) number of comments posted/user (in 1 week)[b]	3.4 (3.9)	2.0 (2.4)	5.4 (10)	0.9 (1.8)	2.2 (4.5)
Mean (SD) number of chat entries/user (in 1 week)[b]	3.9 (5.3)	5.6 (6.1)	8.8 (9.3)	4.0 (4.4)	5.0 (6.0)

[a] The website did not have an automatic time-out feature, as we wanted to encourage users to keep it open on their desktops for periodic daily access. Given that some users remained logged on for a number of hours at a time, in order to estimate accurate "active" usage times, we truncated all logged-in times at 30 minutes after the last "activity" (including any change made, or comment or blog posted). In cases where two activities were separated by >60 minutes, we truncated logged-in times at 30 minutes after the first activity and 30 minutes before the second activity (for a maximum total "active" usage time of 60 minutes between activities).

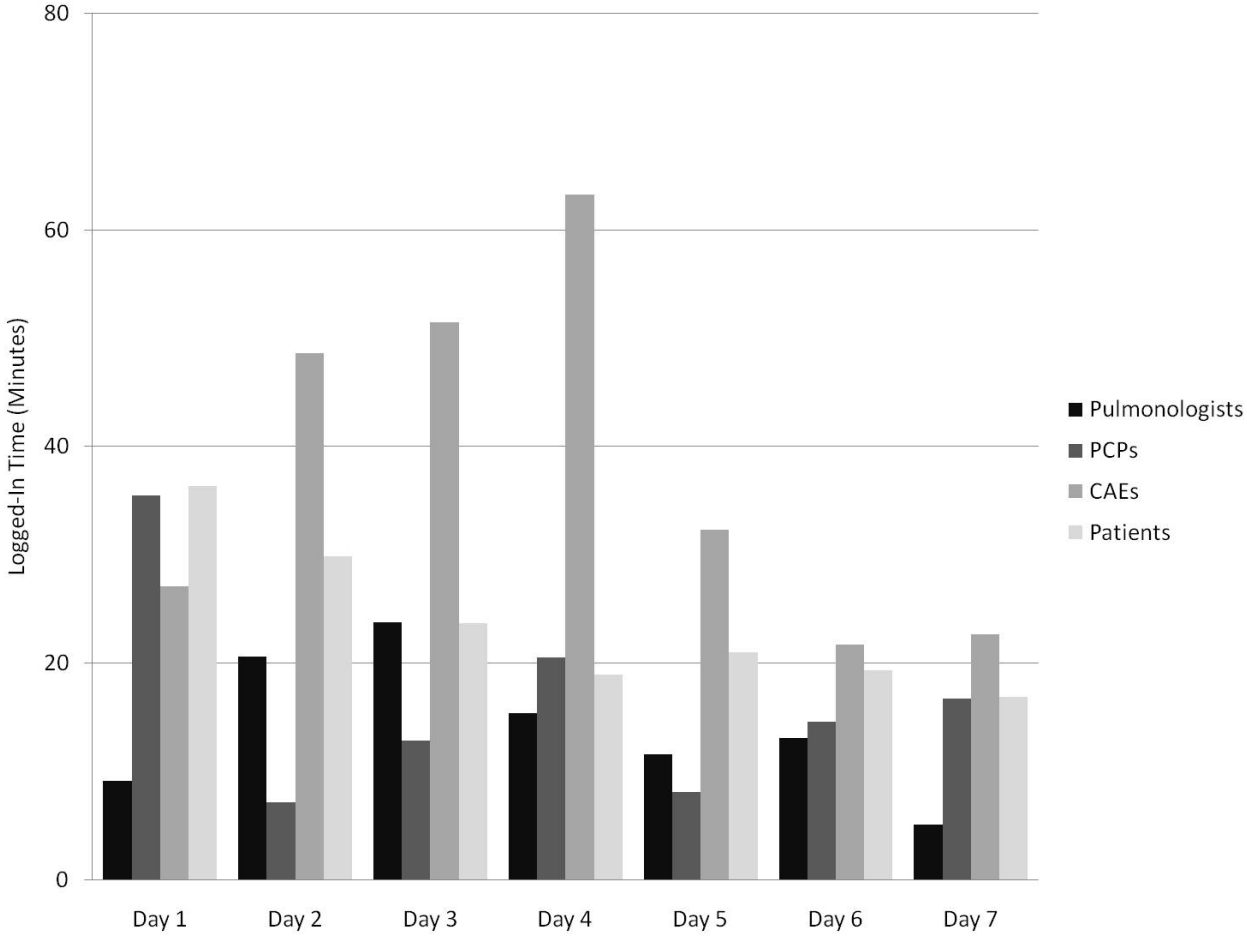
[b] Chat and comment features were used by 28/35 (80%) and 16/35 (46%) participants, respectively.

Temporal trends for site usage and for changes made (by user type) are shown in Figure 2 and Figure 3, respectively. In Figure 2, each bar represents the mean number of minutes that each user was logged in to the system on each day, by user type, and by day of the wiki process. In Figure 3, each bar represents the mean number of changes that each user made to the AAP on each day, by user type, and by day of the wiki process. In both figures, averages were based on all users who had access to the site, and data from all 3 wiki weeks were averaged. A similar decreasing trend for usage (Figure 4) and changes made throughout the week (Figure 5) was seen in each of the 3 wiki weeks. In Figure 4, each bar represents the mean number of minutes that each user was logged in to the system on each day, by wiki week, and by day of the wiki process. In Figure 5, each bar represents the mean number of changes that each user made to the AAP on each day, by wiki week, and by day of the wiki process. In both figures, averages were based on all users who had access to the site, and data from all participants were averaged.

The final AAP had 153 of 229 (67%) choices in common with the AAP at the end of the first day, and 108 of 229 (47%) of the choices that had been made by the end of the first day were not changed through the rest of the week. A detailed description of the final AAP will be published elsewhere (S Gupta, MSc, MD, et al, unpublished data, 2011).

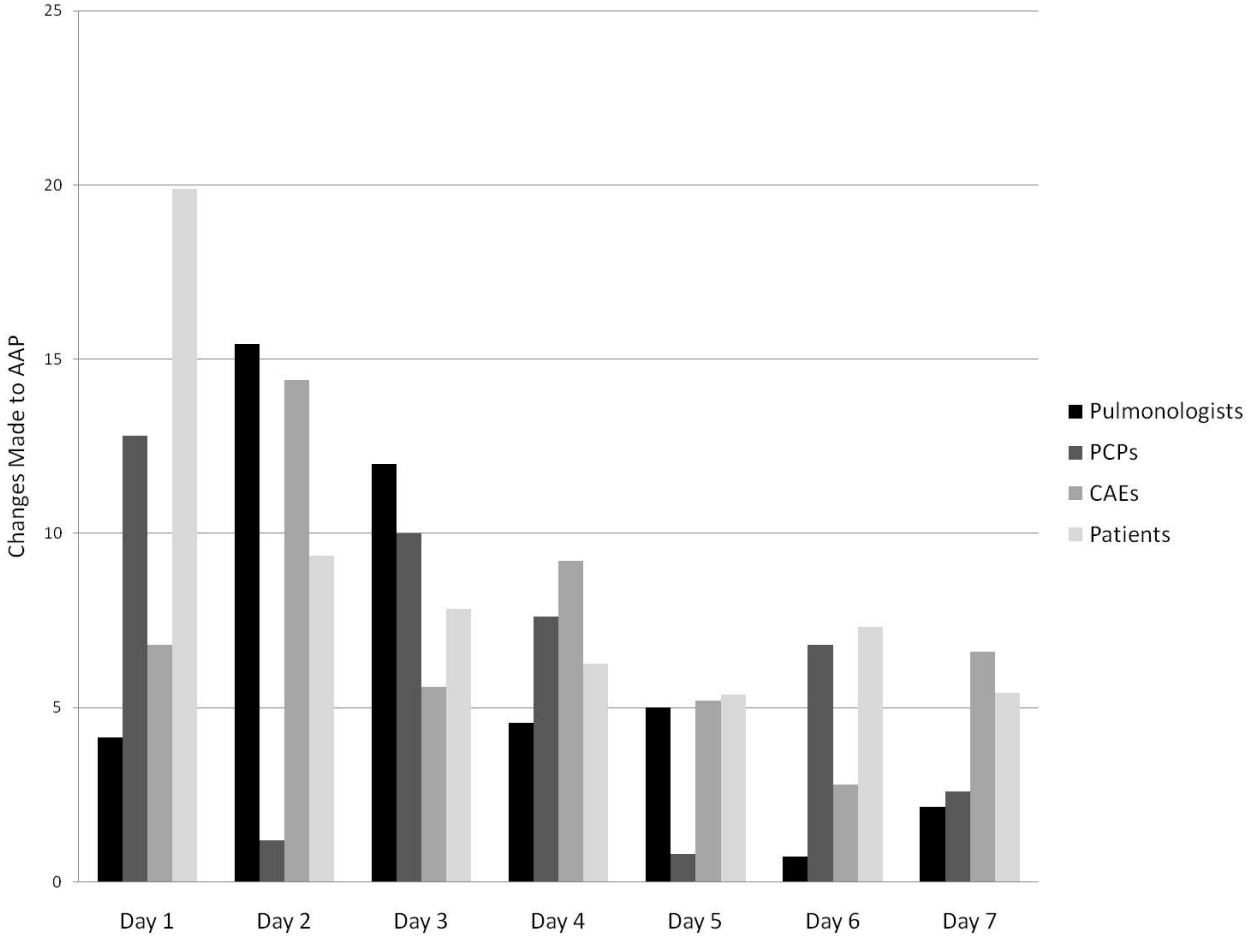
XSL•FO
RenderX

Figure 2. Mean logged-in time per participant per day, by participant type (CAE = certified asthma educator; PCP = primary care physician).



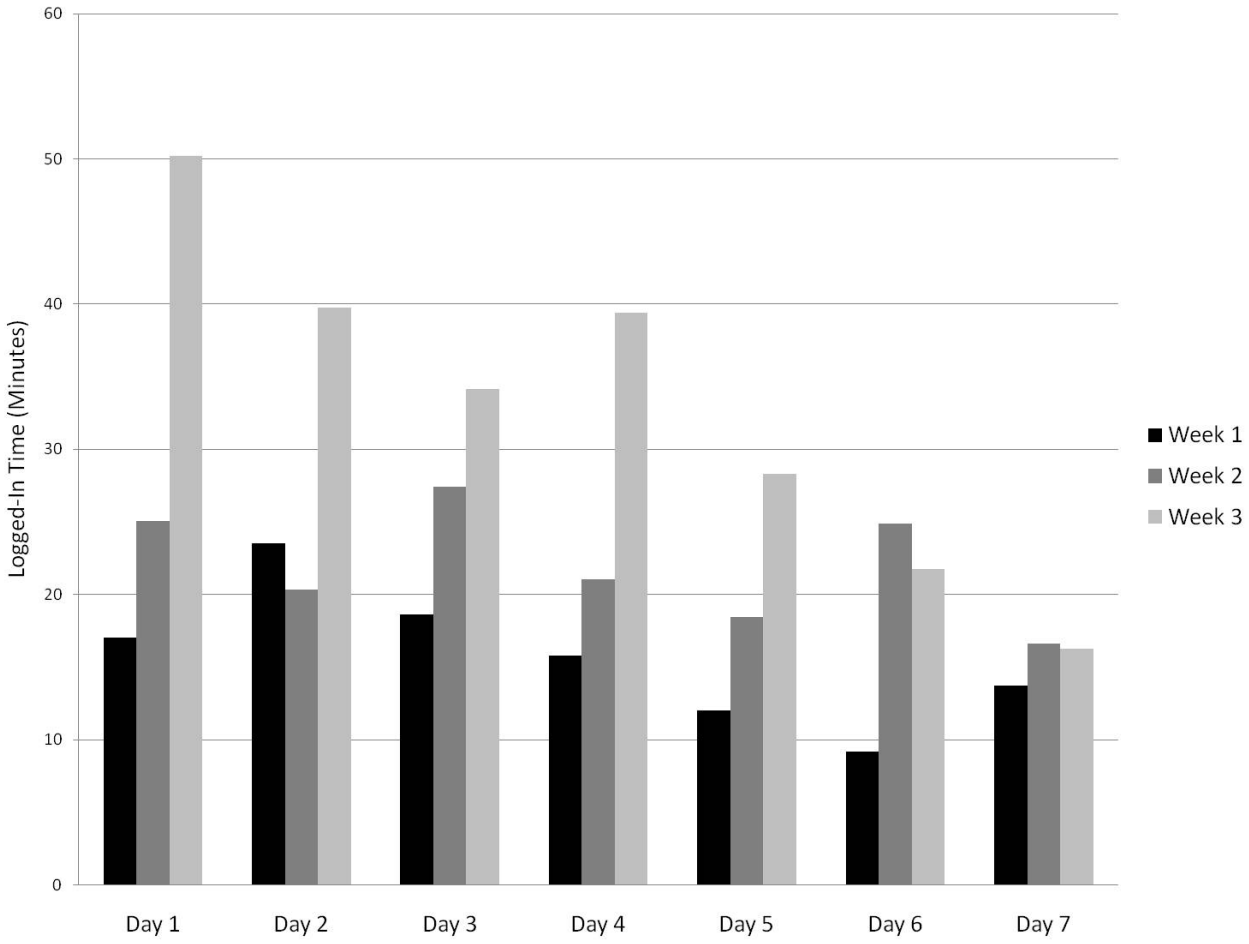
XSL•FO
RenderX

Figure 3. Mean changes made to the wiki-based asthma action plan (AAP) template per participant per day, by participant type (CAE = certified asthma educator; PCP = primary care physician).



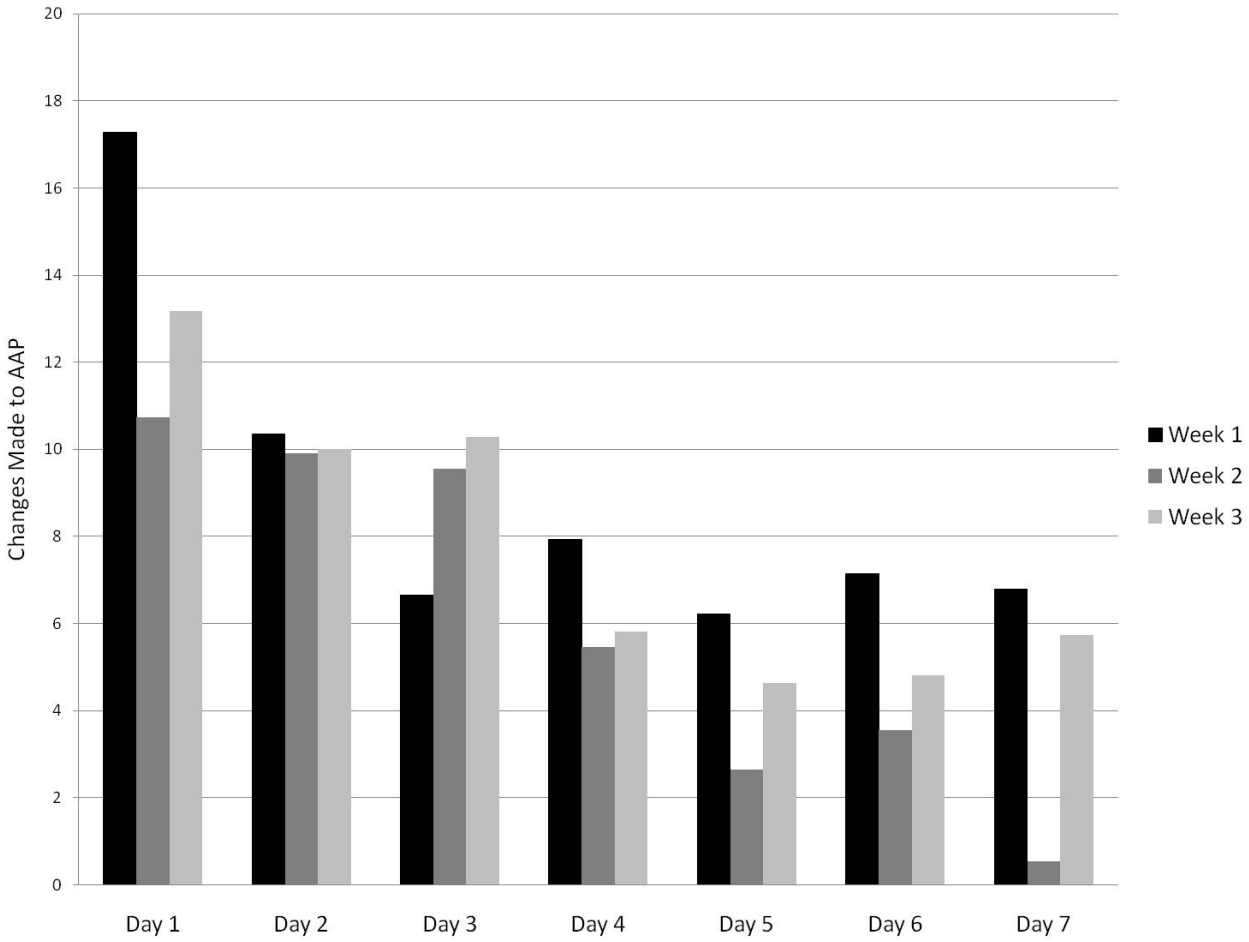
XSL•FO
RenderX

Figure 4. Mean logged-in time per participant per day, by week.



XSL•FO
RenderX

Figure 5. Mean changes made to the wiki-based asthma action plan (AAP) template per participant per day, by week.



User Perceptions

Participant Likert scale responses are summarized in Figure 6, Figure 7, and Figure 8. Of 25 participants, 11 (44%) indicated that they changed their minds about one or more issue(s) based on other participants' preferences.

Reported barriers to tool use included time constraints, difficulties with the login process, no access over work networks, and software bugs. Additional challenges included redundant choices, limited content choices, "information overload," limited amount of space in the AAP, small size of the comment box (requiring frequent scrolling), lack of participant accountability for changes made, difficulty explaining one's point of view through an online chat, and technical challenges understanding site functions. Reported advantages of the system included tool accessibility, broad recruitment, the wide range of available format and content options, tool responsiveness, ability to communicate with peers, and power balance between users, enabling participation by "shy" and "insecure" participants, and averting potentially unpleasant social dynamics. The mean SUS score was 75.9 (SD 19.6).

XSL•FO
RenderX

Figure 6. Specific features of the wiki tool and wiki process. Responses were entered on a 5-point Likert scale labeled as follows: 1, disagree; 3, neutral; and 5, agree. For the purposes of this figure, scores of 1 and 2 were considered "disagree," and 4 and 5 were considered "agree." Each bar demonstrates the proportion of participants with each response, for each statement. This includes 35 participants (5 certified asthma educators, 5 primary care physicians, 6 pulmonologists, and 19 patients with asthma). (AAP = asthma action plan).

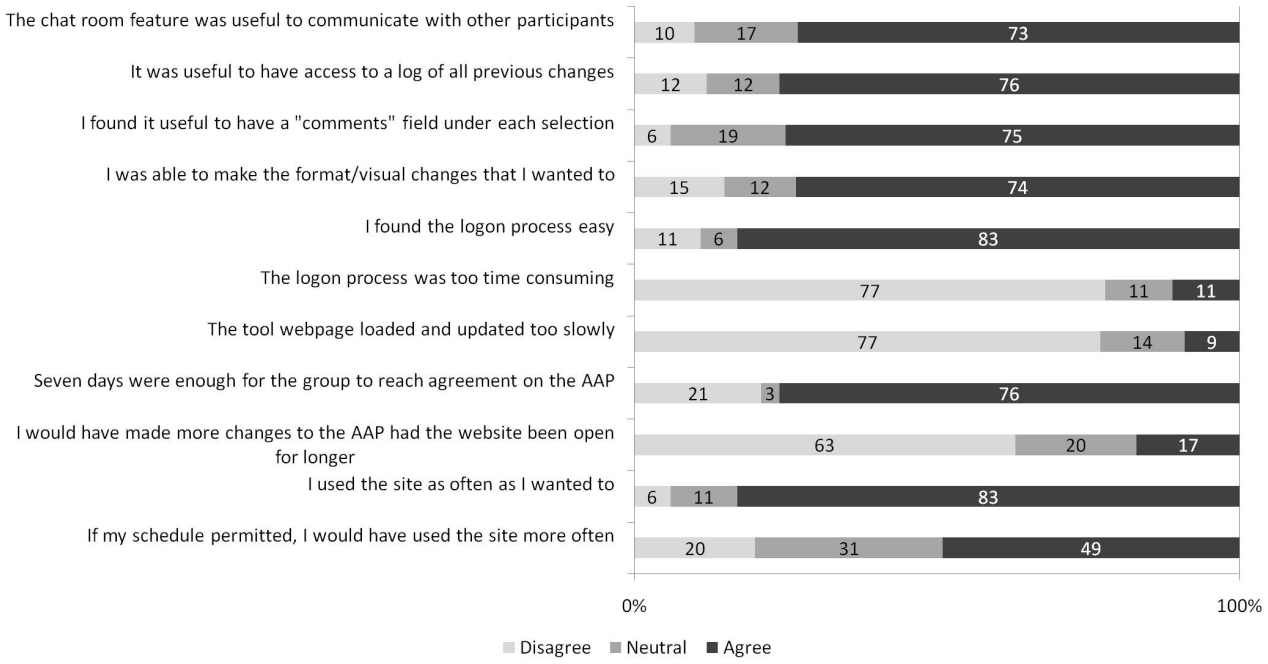
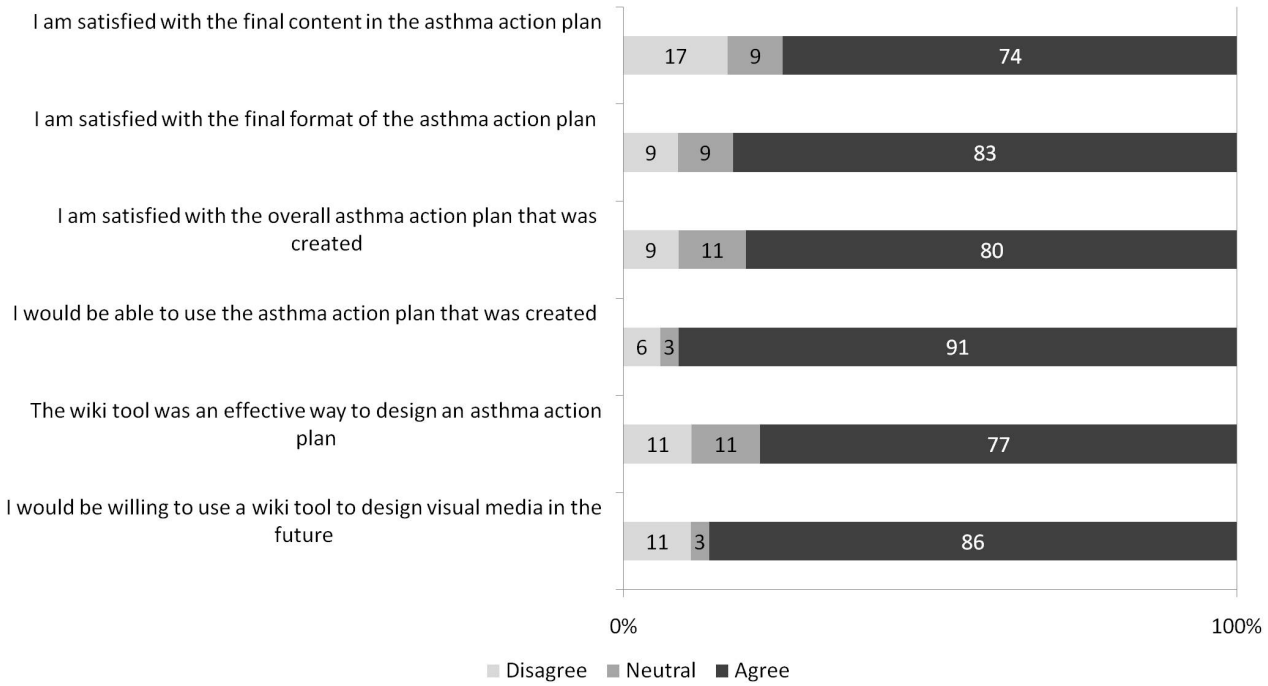
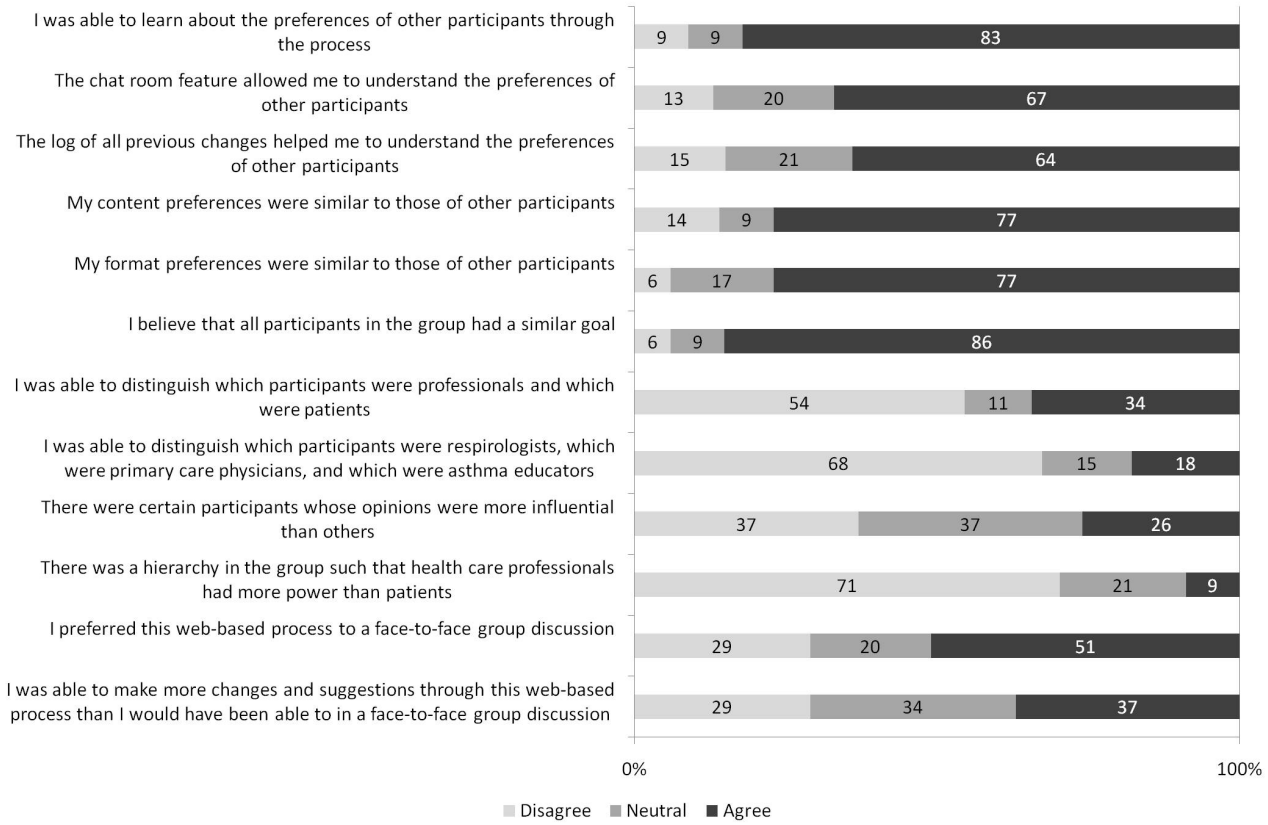


Figure 7. Overall asthma action plan (AAP) and the wiki process. Responses were entered on a 5-point Likert scale labeled as follows: 1, disagree; 3, neutral; and 5, agree. For the purposes of this figure, scores of 1 and 2 were considered "disagree," and 4 and 5 were considered "agree." Each bar demonstrates the proportion of participants with each response, for each statement. This includes 35 participants (5 certified asthma educators, 5 primary care physicians, 6 pulmonologists, and 19 patients with asthma).



XSL•FO
RenderX

Figure 8. Participant interactions. Responses were entered on a 5-point Likert scale labeled as follows: 1, disagree; 3, neutral; and 5, agree. For the purposes of this figure, scores of 1 and 2 were considered "disagree," and 4 and 5 were considered "agree." Each bar demonstrates the proportion of participants with each response, for each statement. This includes 35 participants (5 certified asthma educators, 5 primary care physicians, 6 pulmonologists, and 19 patients with asthma).



Discussion

We developed a novel technique to achieve multiple stakeholder consensus and design, and tested it in the development of an AAP.

Principal Results

Nearly all measures exceeded our expert criteria for an effective technique. Most participants used the tool daily, and each user type actively used the tool for an average of 23 minutes per day. The mean SUS score of 75.9 (SD 19.6) was higher than that in the development stage, indicating system improvement. This score falls in the third quartile of usability scores for other types of tools, indicating good to excellent usability, and is within the range associated with a high chance of real-world user acceptability [25,26]. The data from the SUS can be triangulated with data from the exit questionnaires, which demonstrated greater than 70% user agreement with all usability-related statements, indicating that participants favored its usability highly. Finally, exit questionnaires demonstrated 75% or greater user agreement with nearly all statements relating to satisfaction with the final AAP.

Comparison With Existing Techniques

Conventional focus group or consensus techniques provide limited ability both to measure preferences for and to develop consensus around document aesthetics [4]. Our method has met this challenge; 23.4% (453/1937) of changes related to the format of the AAP, 74% (26/35) of users were able to make the format or visual changes that they wanted to, and 83% (29/53) of participants were satisfied with the AAP's final format. Our process also fulfils the recommended criteria for media design: (1) suitability for all stages of the design process (which can be achieved through iterative wiki design stages), (2) flexibility to adapt to the varying requirements of the design process (which can be achieved through iterative changes to options offered in the wiki tool), and (3) presentation of visual information in a format that inspires users [4]. A caveat is that options in the wiki site must be predetermined, possibly limiting user creativity [4]. An alternative would be to offer users the ability to enter "free text" for tool content, such as in a conventional wiki, although this approach could not easily be applied to format options.

Our method was logistically simpler than other techniques for achieving consensus. Although we used moderators in the development process and in training sessions, unlike in the NGT, consensus development conference, and focus groups, the wiki process does not require a moderator. This method eliminates the task of finding a qualified moderator and associated costs [6], and averts potential pitfalls including poor facilitation, undue influence on participants, and minimization of certain participants' views [27]. A caveat is that site usage was variable and 2 users did not make any changes after logging in; moderators could serve to encourage both universal and more equal participation.

XSL•FO
RenderX

By eliminating the need for participants to meet, we limited organizational and recruitment challenges, and costs incurred with in-person consensus techniques. Our method also enabled international representation, minimizing the geographical bias seen with in-person techniques, and at no incremental cost [9,13]. Another advantage to this method is that preferences are not fixed over time, and attitude change and idea generation may have been enhanced by the 7-day period for interaction, compared with a conventional single in-person meeting [15]. In-person techniques are usually limited to between 5 and 10 participants due to difficulty in coordinating schedules, cost of accommodation, and concerns about group function, with larger sizes favoring unequal participation [9,27]. Our method successfully accommodated groups of 14 participants. Although there is little empirical evidence regarding the effect of the number of participants on the validity of consensus processes, larger groups likely increase the reliability of group judgment [9].

Finally, our method addressed certain challenges arising from complex group dynamics experienced with in-person techniques. Social impact theory suggests that group judgments have a strong influence on individual decision making [9]. In face-to-face consensus processes, participants often define subgroups by stakeholder types. They may be pressured to conform to their own group, and consensus building can be inhibited by intergroup prejudices and stereotypes [9]. Furthermore, the status of individual participants affects their influence on other participants, and status hierarchies—both between professionals and patients, and within professional groups—are likely to emerge in face-to-face meetings [9,28]. These hierarchical structures limit the willingness of some participants to contribute openly [13]. The wiki method minimizes any group or individual influence by anonymizing stakeholders and by eliminating verbal communication, which is both a source and an indicator of status within groups [28]. Although certain participant group identities were revealed through chat and comment entries, only 9% (3/35) of participants perceived a power differential between users. Overall, most participants reported successfully learning about the preferences of others and believed that everyone had a similar goal; 37% (13/35) were also able to make contributions that they felt they could not have made in a face-to-face forum. However, it is possible that HCPs prefer the hierarchical power differential that they enjoy in face-to-face forums. This may partly explain the dichotomous responses to whether participants preferred the Web-based process to a face-to-face discussion, and whether they felt that the Web-based process enabled them to effectively make more changes (Figure 8).

The Delphi method shares the resource, time, and recruitment advantages of the wiki method and avoids concerns related to group functioning. However, in contrast to the wiki method, the Delphi method does not capture the important synergistic effect of participant interaction on the development of ideas [4], lacks a mechanism for conflict resolution in areas of disagreement (enabled by chat room discussions, comment fields, and the online vote feature in the wiki method), and does not identify the reasons for disagreements (enabled by qualitative analysis of chat room discussions and comment fields in the wiki method) [9]. The online Delphi method shares the same platform as the wiki method (the Internet) and has been shown to be more time and cost efficient than the traditional Delphi method [10,11], but shares the limitations described above. In addition, Web content is limited to text, and this method does not enable measurement of group preferences or group collaborative development around document aesthetics [10].

Limitations

We noted some logistical and technical difficulties. A moderate proportion of the targeted sample had no or limited participation, in some cases due to unexpected changes in personal circumstances, but in most cases due to software access difficulties caused by firewalls. Although full-time representation from each user type was maintained in all 3 wiki weeks in this study, differential dropout rates between user types can threaten the validity of the process and should be addressed in future studies. Researchers should anticipate such losses in setting recruitment targets, instruct users to test all required websites on all computers that they intend to use during the process, and verify site functionality with users before the process begins.

Users also struggled with software instructions, as demonstrated by the large number of failed logins, likely because Safari was not their default browser. This did not have a significant impact on usage, as all but 1 participant successfully logged on. However, this demonstrated the importance of 24-hour technical support in enabling this process. Future studies of the wiki method should emphasize use of the correct browser at the orientation stage and consider building cross-browser compatibility.

Although the wiki process averted the costs of hiring a moderator and organizational and travel costs incurred in in-person techniques, development process costs were considerable. These include costs for software development, moderated in-person tool testing, and tool revisions based on user feedback. Technical support and analysis of wiki session data are additional costs.

There are several caveats regarding the effect of the wiki process on group functioning. Attitude change is an important part of consensus building [9]. This requires persuasive ability, which can be influenced by visual cues (eg, facial expressions), and paralinguistic cues (eg, voice quality), and depends on the credibility, trustworthiness, and likeability of the communicator, all of which may be better transmitted through in-person interactions [9]. Elements such as tone of voice, and facial and body expressions are useful “human” cues that are lost in the wiki method [4]. Although the effect of online communication (either synchronous or asynchronous) on individual decision making is unknown, 44% (11/25) of participants indicated that they changed their minds about one or more issue(s) based on the preferences of their peers. However, a minority believed that they were able to make more changes in the Web-based tool than they could have in a face-to-face forum. This may also relate to the fact that user options were limited to those offered in drop-down menus, as opposed to a theoretically unlimited number of options available in a face-to-face discussion.

XSL•FO
RenderX

Motivation to participate in this process may be similar to motivations that have made Wikipedia one of the most visited Internet sites in the world—the pleasure, validation, and sense of ownership that users derive from seeing their personal edits and contributions [16]. Although overall user engagement was strong, 2 users did not make any changes. This "social loafing," whereby certain group members leave the bulk of the work to others [9], may be exacerbated by the anonymity in the wiki process, compared with face-to-face processes, where social loafers risk embarrassment. Anonymity might also facilitate a contrarian or destructive contribution pattern, and although this is unlikely in a carefully recruited group, lack of accountability was cited as a disadvantage by 1 user.

Organizational psychological research suggests that participants' initially expressed opinions may disproportionately influence consensus group decisions [9]. The wiki process is particularly susceptible to this bias, as it begins with a "blank slate" that the first few participants alter to create the first recognizable form of the tool, which is then edited by others. We noted that 67% (153/229) of the choices in the final AAP had been made at the end of the first day, and 47% (108/229) of the choices that had been made at the end of the first day were never changed throughout the week. Conversely, the wiki process is susceptible to a single participant or a small group of participants making substantial changes just before the process end time, threatening the collaborative or consensus nature of the outcome. Remarkably, this did not occur in any of the 3 wiki weeks. Furthermore, as Figure 3 shows, the mean number of changes made per day trended downward as the week progressed, with the fewest changes made in the last 3 days. This may reflect progressively increasing user satisfaction with the developed product.

Conclusions

We developed a unique wiki-inspired process for collaborative design and consensus building and applied it in the development of an AAP. The name "WikiBuild" was chosen to capture the essence of this technology—a wiki-inspired tool designed to enable users to collaboratively build a multitude of different products. Potential uses of our method are broad, and include development of both medical and nonmedical tools and products. Commercial uses may include development of marketing material by members of the target consumer group itself, or codevelopment by designers and target consumers. Future studies should address software- and hardware-related technical challenges and questions about the dynamics and functioning of virtual focus groups. Novel variations of our study design should also be considered, such as testing larger wiki groups, or running separate wiki processes in different stakeholder types in order to explore differences between stakeholder preferences.

Acknowledgments

This work was supported by the Physicians' Services Incorporated Foundation and the Bell University Labs. The funders played no role in the design of the study, the collection, analysis, and interpretation of data, the writing of the manuscript, or the decision to submit the manuscript for publication. We would like to acknowledge Susan Hall, the research coordinator for this study. We would like to thank Dr. Francine M. Ducharme, Dr. Andreas Laupacis, Dr. Diane Lougheed, Dr. Ann Mckibbon, Ms. Carole Madeley, and Ms. Maria Wowk for their assistance in planning and executing this study. We would also like to thank all study participants for their valuable time.

Conflicts of Interest

None declared

Multimedia Appendix 1

Development stage inclusion criteria.

[PDF file (Adobe PDF File), 58 KB - jmir_v13i4e108_app1.pdf]

Multimedia Appendix 2

Video demonstrating the evolution of the AAP over the first wiki week.

[MP4 File (MP4 Video), 25MB - jmir_v13i4e108_app2.mp4]

References

1. Gibson PG, Powell H, Coughlan J, Wilson AJ, Abramson M, Haywood P, et al. Self-management education and regular practitioner review for adults with asthma. Cochrane Database Syst Rev 2003(1):CD001117. [doi: 10.1002/14651858.CD001117] [Medline: 12535399]
2. Douglass J, Aroni R, Goeman D, Stewart K, Sawyer S, Thien F, et al. A qualitative study of action plans for asthma. BMJ 2002 Apr 27;324(7344):1003-1005 [FREE Full text] [Medline: 11976240]
3. FitzGerald JM, Boulet LP, McIvor RA, Zimmerman S, Chapman KR. Asthma control in Canada remains suboptimal: the Reality of Asthma Control (TRAC) study. Can Respir J 2006;13(5):253-259 [FREE Full text] [Medline: 16896426]

XSL•FO
RenderX

4. Bruseberg A, McDonagh-Philp D. Focus groups to support the industrial/product designer: a review based on current literature and designers' feedback. Appl Ergon 2002 Jan;33(1):27-38. [Medline: 11827136]
5. Keown K, Van Eerd D, Irvin E. Stakeholder engagement opportunities in systematic reviews: knowledge transfer for policy and practice. J Contin Educ Health Prof 2008;28(2):67-72. [doi: 10.1002/chp.159] [Medline: 18521874]
6. Evans KE, Klinger T. Obstacles to bottom-up implementation of marine ecosystem management. Conserv Biol 2008 Oct;22(5):1135-1143. [doi: 10.1111/j.1523-1739.2008.01056.x] [Medline: 18954339]
7. AGREE Collaboration. Health Care Evaluation Unit, St George's Hospital Medical School, London, UK. 2004 Mar. Appraisal of Guidelines Research and Evaluation (AGREE) Instrument URL: http://www.agreecollaboration.org/instrument/ [accessed 2011-09-30] [WebCite Cache ID 625hY8RdM]
8. Vastine A, Gittelsohn J, Ethelbah B, Anliker J, Caballero B. Formative research and stakeholder participation in intervention development. Am J Health Behav 2005;29(1):57-69. [Medline: 15604050]
9. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess 1998;2(3):i-iv, 1 [FREE Full text] [Medline: 9561895]
10. Deshpande AM, Shiffman RN, Nadkarni PM. Metadata-driven Delphi rating on the Internet. Comput Methods Programs Biomed 2005 Jan;77(1):49-56. [doi: 10.1016/j.cmpb.2004.05.006] [Medline: 15639709]
11. Snyder-Halpern R, Thompson CB, Schaffer J. Comparison of mailed vs. Internet applications of the Delphi technique in clinical informatics research. Proc AMIA Symp 2000:809-813. [Medline: 11079996]
12. Reingold S, Kulstad E. Impact of human factor design on the use of order sets in the treatment of congestive heart failure. Acad Emerg Med 2007 Nov;14(11):1097-1105. [doi: 10.1197/j.aem.2007.05.006] [Medline: 17967968]
13. Mansell I, Bennett G, Northway R, Mead D, Moseley L. The learning curve: the advantages and disadvantages in the use of focus groups as a method of data collection. Nurse Res 2004;11(4):79-88. [Medline: 15227901]
14. Twohig PL, Putnam W. Group interviews in primary care research: advancing the state of the art or ritualized research? Fam Pract 2002 Jun;19(3):278-284 [FREE Full text] [Medline: 11978719]
15. Tompkins EL, Few R, Brown K. Scenario-based stakeholder engagement: incorporating stakeholders preferences into coastal planning for climate change. J Environ Manage 2008 Sep;88(4):1580-1592. [doi: 10.1016/j.jenvman.2007.07.025] [Medline: 17923247]
16. Agha R. Introducing Wikisurgery.com: the blueprint for a surgical architecture of participation. Int J Surg 2006;4(3):140-143. [doi: 10.1016/j.ijsu.2006.07.002] [Medline: 17462335]
17. Seebregts CJ, Mamlin BW, Biondich PG, Fraser HS, Wolfe BA, Jazayeri D, OpenMRS Implementers Network. The OpenMRS Implementers Network. Int J Med Inform 2009 Nov;78(11):711-720. [doi: 10.1016/j.ijmedinf.2008.09.005] [Medline: 19157968]
18. Salzberg SL. Genome re-annotation: a wiki solution? Genome Biol 2007;8(1):102 [FREE Full text] [doi: 10.1186/gb-2007-8-1-102] [Medline: 17274839]
19. Giles J. Key biology databases go wiki. Nature 2007 Feb 15;445(7129):691. [doi: 10.1038/445691a] [Medline: 17301755]
20. Lau AS. Implementation of an onto-wiki toolkit using web services to improve the efficiency and effectiveness of medical ontology co-authoring and analysis. Inform Health Soc Care 2009 Jan;34(1):73-80. [doi: 10.1080/17538150902779543] [Medline: 19306201]
21. Streeter JL, Lu MT, Rybicki FJ. Informatics in radiology: RadiologyWiki.org: the free radiology resource that anyone can edit. Radiographics 2007;27(4):1193-1200 [FREE Full text] [doi: 10.1148/rg.274065090] [Medline: 17620476]
22. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23. [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
23. Brooke J. SUS: a ‘quick and dirty’ usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry. London, UK: Taylor & Francis; 1996:189-194.
24. Wikipedia. 2011 Jun 16. Wiki URL: http://en.wikipedia.org/wiki/Wiki [accessed 2011-06-16] [WebCite Cache ID 5zUuZx48C]
25. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008;24(6):574-594. [doi: 10.1080/10447310802205776]
26. Kastner M, Lottridge D, Marquez C, Newton D, Straus SE. Usability evaluation of a clinical decision support tool for osteoporosis disease management. Implement Sci 2010;5:96 [FREE Full text] [doi: 10.1186/1748-5908-5-96] [Medline: 21143978]
27. Pagliari C, Grimshaw J, Eccles M. The potential influence of small group processes on guideline development. J Eval Clin Pract 2001 May;7(2):165-173. [Medline: 11489041]
28. Pagliari C, Grimshaw J. Impact of group structure and process on multidisciplinary evidence-based guideline development: an observational study. J Eval Clin Pract 2002 May;8(2):145-153. [Medline: 12180363]

Abbreviations

AAP: asthma action plan

XSL•FO
RenderX

CAE: certified asthma educator
HCP: health care professional
NGT: nominal group technique
PCP: primary care physician
SUS: System Usability Scale

Edited by G Eysenbach; submitted 16.04.11; peer-reviewed by P Maude, P Archambault; comments to author 13.05.11; revised version received 09.07.11; accepted 09.07.11; published 08.12.11.

Please cite as:
Gupta S, Wan FT, Newton D, Bhattacharyya OK, Chignell MH, Straus SE
WikiBuild: A New Online Collaboration Process For Multistakeholder Tool Development and Consensus Building
J Med Internet Res 2011;13(4):e108
URL: http://www.jmir.org/2011/4/e108/
doi:10.2196/jmir.1833
PMID:22155694

©Samir Gupta, Flora T Wan, David Newton, Onil K Bhattacharyya, Mark H Chignell, Sharon E Straus. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Beyond Readability: Investigating Coherence of Clinical Text for Consumers

Catherine Arnott Smith[1], AMLS, MS, MA, PhD; Scott Hetzel[2], MS; Prudence Dalrymple[3], MS, PhD; Alla Keselman[4], MA, PhD

[1]School of Library and Information Studies, University of Wisconsin-Madison, Madison, WI, United States

[2]Biostatistics and Medical Informatics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States

[3]Institute for Healthcare Informatics, College of Information Science and Technology, Drexel University, Philadelphia, PA, United States

[4]Division of Specialized Information Services, National Library of Medicine, Bethesda, MD, United States

Corresponding Author:
Catherine Arnott Smith, AMLS, MS, MA, PhD
School of Library and Information Studies
University of Wisconsin-Madison
600 N Park Street
Madison, WI, 53706
United States
Phone: 1 608 890 1334
Fax: 1 608 263 4849
Email: casmith24@wisc.edu

Abstract

Background: A basic tenet of consumer health informatics is that understandable health resources empower the public. Text comprehension holds great promise for helping to characterize consumer problems in understanding health texts. The need for efficient ways to assess consumer-oriented health texts and the availability of computationally supported tools led us to explore the effect of various text characteristics on readers' understanding of health texts, as well as to develop novel approaches to assessing these characteristics.

Objective: The goal of this study was to compare the impact of two different approaches to enhancing readability, and three interventions, on individuals' comprehension of short, complex passages of health text.

Methods: Participants were 80 university staff, faculty, or students. Each participant was asked to "retell" the content of two health texts: one a clinical trial in the domain of diabetes mellitus, and the other typical Visit Notes. These texts were transformed for the intervention arms of the study. Two interventions provided terminology support via (1) standard dictionary or (2) contextualized vocabulary definitions. The third intervention provided coherence improvement. We assessed participants' comprehension of the clinical texts through propositional analysis, an open-ended questionnaire, and analysis of the number of errors made.

Results: For the clinical trial text, the effect of text condition was not significant in any of the comparisons, suggesting no differences in recall, despite the varying levels of support (P = .84). For the Visit Note, however, the difference in the median total propositions recalled between the Coherent and the (Original + Dictionary) conditions was significant (P = .04). This suggests that participants in the Coherent condition recalled more of the original Visit Notes content than did participants in the Original and the Dictionary conditions combined. However, no difference was seen between (Original + Dictionary) and Vocabulary (P = .36) nor Coherent and Vocabulary (P = .62). No statistically significant effect of any document transformation was found either in the open-ended questionnaire (clinical trial: P = .86, Visit Note: P = .20) or in the error rate (clinical trial: P = .47, Visit Note: P = .25). However, post hoc power analysis suggested that increasing the sample size by approximately 6 participants per condition would result in a significant difference for the Visit Note, but not for the clinical trial text.

Conclusions: Statistically, the results of this study attest that improving coherence has a small effect on consumer comprehension of clinical text, but the task is extremely labor intensive and not scalable. Further research is needed using texts from more diverse clinical domains and more heterogeneous participants, including actual patients. Since comprehensibility of clinical text appears difficult to automate, informatics support tools may most productively support the health care professionals tasked with making clinical information understandable to patients.

XSL•FO
RenderX

(J Med Internet Res 2011;13(4):e104) doi:10.2196/jmir.1842

KEYWORDS

Health literacy; comprehension; vocabulary; patients; language tests; retention (psychology)

Introduction

A basic tenet of consumer health informatics is that understandable health resources empower the public by increasing knowledge and improving decision making [1]. Research indicates that most laypeople have difficulty comprehending medical documents, especially those that describe complex information pertaining to clinical research (for example, [2]). For example, numerous studies suggest that most patients, especially those with lower income levels and literacy skills, have difficulty reading and comprehending informed consent documents [3]. Poor understanding of health information thus has an impact on public health. This has prompted many research efforts to close the gap between the difficulty level of documents and readers' literacy by improving the readability of health-related materials.

Readability itself is a concept drawn from kindergarten to grade 12 education, in which field research findings suggest that readers' ability to comprehend a passage decreases as the number of "difficult" words (that is, words unfamiliar to the average reader) increases. Typically, readability measures are derived from sentence and word length. Substituting simpler, more familiar vocabulary improves readability in texts whose subject matter addresses general knowledge domains. (For a comprehensive review, see [4]).

Attempts to lower the readability level of health materials, usually to grades 7–9, have successfully employed these techniques, but health information presents additional challenges. Health information not only contains unfamiliar and difficult words, but also abounds with complex concepts such as those related to physiology and pharmacology. In addition, understanding health materials requires readers to make inferences that access a body of specialized knowledge supporting the information. Experts possess this specialized body of knowledge and so are able to make inferences, but even highly educated nonspecialists may not be able to make those inferences because they lack the necessary training and experience.

Literature Review

Patients and Clinical Language

Medical terminology has long been recognized as a specialized language that is acquired through education and clinical practice [5]. For decades, medical terminology has been cited by physicians as a significant concern about patients' possible misinterpretation of medical record content [6-10]. Much of this literature derives from early controversies over patient access to records—first, in psychiatry, proposed initially as an entirely theoretical construct by Westin [11], and then, in the early years of the British Access to Health Records Act, which in 1990 first gave British patients access to their medical information "held in manual form" [8].

Surprisingly, given the early concerns expressed about patient understanding of medical terminology, few studies published since the mid-1990s have examined the problem. Only Tomkins et al [12] examined patients' comprehension of physician letters. Far more common is the assertion that medical terminology is a nail, and customized patient education materials the appropriate hammer. For example, Nijland et al [13] found that terminology was a barrier to usability in patient self-care applications, concluding that "Self-care support applications should match the vocabulary of the users and the language of the medical systems."

The medical terminology problem is compounded by the consumer health vocabulary problem: that the everyday language used by consumers to describe diseases and treatments is a hybrid of specialized terms and common words that are part of general spoken vocabularies (see Keselman et al [14] for a discussion of research in this area to date). Consumer terms are also characterized by levels of granularity and specificity different from their medical counterparts. For example, anatomical words such as "blood" or "brain" usually suffice for lay discussion of physiology or pathology, while the specialist requires terms that describe much smaller, more specific aspects of the blood or the brain. For these reasons, any approach to vocabulary simplification is challenging, particularly for analysis by reading level. Words that contain many syllables, such as "hospitalization," contribute to a higher calculated reading level for that document, yet the term hospitalization is easily understood by laypeople who know what a hospital is; conversely, short and simple words such as "gene" or "immune" are terms referencing complex entities and processes.

Carefully designed entry vocabularies may be able to serve as bridges between terms and concepts used by different user communities. This idea was the impetus for development of the Unified Medical Language System, which focuses on the numerous sublanguages of health care. The idea that consumers constitute a user community of their own, however, is more recent (see [15] for a review), and details of how such an entry vocabulary can be used in real-world implementations are lacking. Zeng and Tse [16] argued in opposition to Patrick et al [17] that simply providing users with a list of medical terms or a dictionary would not solve the terminology problem for informatics applications. More recently, Leroy and Miller [18] found some evidence supporting Zeng and Tse. This reading comprehension study investigated the effect of automatically generated health topics overviews (HTOs). These HTOs, described as functioning "much as a table of contents," were overviews, not dictionaries, but like dictionaries were intended to function as information assists. Leroy and Miller found that "vulnerable" consumers—those identified as having low health literacy or high stress—were found to rely on the HTO even more than they did on the text that the HTO summarized—so much so that vulnerable consumers performed more poorly when an HTO was not available.

XSL•FO
RenderX

Dictionaries can be considered a good starting point for language bridges, since they contain definitions presumed to be standard and thus common across communities. However, the sociocultural dimensions of term variability require more depth and breadth of expression than any dictionary or glossary can effectively provide without a great deal of customization; Velardi and colleagues have commented on the resource intensiveness of the interactive glossary process itself [19]. Indeed, the literature of medical informatics is largely silent on this question, since the usefulness of dictionaries or glossaries as terminology support in health informatics is an untested assumption, with one exception: Diefenbach and Butz [20] constructed a virtual health center for use in educating patients with prostate cancer. The library "room" in this health center used a glossary in which some medical terms were hyperlinked to short definitions. A focus group of patients and spouses identified the glossary as helpful.

Coherence

Text coherence, a concept from the fields of cognitive psychology and education, refers to the connectedness of ideas in a text, which affects comprehension [21]. A distinction is usually made between *local* and *global* coherence. Local coherence refers to the explicit connection between adjacent clauses and sentences, also referred to as *cohesion*; global coherence refers to the logical organization in which macro-level ideas are presented. [22].Text coherence is the function of both text and reader; an identical text may be perceived as either well or poorly connected, depending on the reader's background knowledge. Extracting meaning from text always requires some inferences, and it is the background knowledge that determines whether the needed inferences are trivial or insurmountable.

When discussed as a feature of a text, coherence usually applies to its "average" reader, or target audience. When it comes to comprehending medical information, laypeople lack the in-depth knowledge of the domain, an expertise that is characteristic of the professional who wrote those texts. Medical documents such as those contained in electronic and personal health records, informed consent forms, and medication instructions are likely to refer to concepts and make connections with which readers may not be familiar.

To support consumer comprehension, we must bridge the gap that exists between writers' and readers' knowledge: between the elaborate background expertise of the writer that serves as a basis for assumptions in the text, and the more modest background familiarity of the reader. In other words, we need to make the text more "coherent," ensuring that its ideas are well connected not only with each other, but with the likely background knowledge of the intended reader; that the number of inferences, or mental leaps, required of the reader moving from one idea to the next is not excessive; and that these inferences are easy to make.

Consider the following statement: "After Jennifer mentioned that her daughter was 'always thirsty,' the doctor asked if she had recently lost weight." A reader with some knowledge about type 1 diabetes will know that thirst and weight loss are both symptoms of diabetes. Such a knowledgeable reader will infer the connection between the two clauses of the sentence, will understand that the word "she" refers to the daughter rather than the mother, and may even anticipate the doctor's next questions. The reader without prior knowledge of diabetes, however, will not be able to make the connection.

Coherent texts ensure that less effort is required for the reader to transition from clause to clause, extracting meaning and building a mental representation of the text. In comprehension research, text memory and mental representations are typically measured in terms of *propositions*. A proposition is the smallest meaningful unit of thought, often consisting of two concepts and a relationship that connects them (antiobiotic_TREATS_infection), or a concept and a modifier (infection_IS_bacterial). Propositions typically correspond to sentence clauses. Not every proposition of the original text is encoded and remembered [23]: concepts and relationships that are connected to the reader's prior knowledge are more likely to be retained.

Reduced comprehension effort is not necessarily always better for all readers. In fact, studies suggest that when readers with strong background read less coherent texts, they are forced into deeper processing, and actually learn more [21,24]. For less knowledgeable readers, however, lack of coherence in the text is detrimental to comprehension and learning. As they lack background knowledge concepts to which they can relate the text, they remember little, and build representations characterized by omissions and errors [25]. Laypeople reading medical documents are likely to fall into the category of less knowledgeable readers, those whose comprehension would benefit from more coherent texts.

Little is known about the coherence of standard medical documents, because research into the comprehensibility of these materials has typically focused on readability. As noted above, however, readability does not ensure coherence. Local coherence is likely to be compromised by the unfamiliar concepts and relationships between them—as in the example given above regarding thirst and weight loss—as well as by general writing style issues, neither specific nor limited to the health domain. Global coherence, additionally, is likely to be compromised if the overall structure of the documents reflects health professionals' rather than lay conception of health and disease.

Studies in cognitive psychology suggest that rewriting texts using explicit coherence principles, rather than writers' intuition, leads to improved comprehension for less knowledgeable readers. McNamara et al [26] analyzed 12 available studies that revised texts to change their coherence (which these authors refer to as cohesion). Principles for improving local coherence typically involve strategies such as the addition of argument overlap (making each sentence repeat the linking word from a previous sentence), the use of sentence connectives, and the rearrangement of clauses so that sentences repeat old ideas before introducing new ones. Improving global coherence involves introducing background concepts; making important references explicit; explaining causal connections between events; adding headers and topic sentences; and clearly linking subtopics to the main topic [21,25,27,28].

Although published studies describe many strategies for improving text coherence, they do not provide specific guidance

XSL•FO
RenderX

for choosing among them. Most studies use a combination of techniques, directed at improving both local and global coherence. Vidal-Abarca et al [27] explicitly compared the effect of local versus global coherence improvements in a history text on the Russian Revolution, and concluded that global, but not local, coherence improvements led to deeper comprehension, as measured by the ability to answer inference questions and focus on main ideas during recall. Vidal-Abarca and colleagues also concluded that the strongest benefits for comprehension were produced by a version with both local and global coherence revisions. To the best of our knowledge, no studies have compared the impact of local versus global comprehension revisions as a function of text difficulty. It is reasonable to expect that global coherence revisions, which target gaps in knowledge, are more essential for texts in knowledge-rich domains, such as history or medicine, than for domains with weaker ties to specialized subject knowledge, such as fiction. At the same time, one should keep in mind that conceptual complexity (and thus global coherence) is not purely a characteristic of a text, but of a match between the text, the knowledge and intention of its authors, and its reader. This makes global coherence editing more art than science, compared with local coherence editing.

While the cognitive psychology literature outlines rather specific principles for improving text coherence, professional writers of health education brochures have a wide range of notions about what it means for the text to be coherent and how coherence can be achieved (Kools et al [29].) Certain specific principles of coherence were overlooked by these writers—for example, the use of sentence connectives to clarify relationships, especially causal relationships, between concepts; and the correct use of word order, to make clear that new information is related to information previously given to the reader. Extending the focus of consumer health comprehension research beyond readability to include coherence is likely to lead to insights about ways to support patients' understanding of medical documents.

Study Goals

The need for efficient ways to assess consumer-oriented health texts, and the availability of computationally supported tools to accomplish these tasks, led us to explore the effect of various text characteristics on readers' understanding of health texts, as well as to develop novel approaches to assessing these characteristics. We were particularly interested in coherence and the complexity of health-specific vocabulary. To explore these issues, we conducted an exploratory study to compare two approaches to improving the readability of health materials. One approach focuses on identifying and explaining difficult words; the other focuses on identifying logical gaps and providing additional texts to facilitate inference, thereby increasing coherence.

The goal of this study was to compare the impact of three interventions on individuals' comprehension of short, complex passages of health text. Two interventions provided terminology support via (1) standard dictionary, or (2) contextualized vocabulary definitions developed specifically for the study. The third intervention provided coherence improvement. The Methods section describes these interventions in detail. We tested the following 4 hypotheses.

H1: Readers' comprehension of a text enhanced by providing standard, off-the-shelf dictionary definitions (hereafter referred to as the Dictionary condition) will be equivalent to their comprehension of the original text (hereafter referred to as the Original condition).

H2: Readers' comprehension of a vocabulary-enhanced text (hereafter referred to as the Vocabulary condition) will be significantly greater than in the Original and Dictionary conditions combined.

H3: Readers' comprehension of a text with improved coherence (hereafter referred to as the Coherent condition) will be significantly greater than in the Original plus Dictionary conditions.

H4: Readers' comprehension of the Coherent condition will be significantly greater than in the Vocabulary condition.

Methods

Participants

A total of 80 people associated with the University of Wisconsin-Madison as staff, faculty, or students participated in the study. Participants were recruited in two cohorts. The first cohort of 40 participants consisted of mixed faculty, staff, graduate, and undergraduate students recruited via campus fliers and newspaper advertising. After we determined that the initial sample size was insufficient to capture the effects of the interventions, we recruited a second cohort, consisting entirely of graduate students in library and information studies, from the University of Wisconsin-Madison School of Library and Information Studies, via an in-class announcement. All participants completed the tasks individually and received $25 bookstore gift cards for participating. The study was approved by the Social Sciences Institutional Review Board of the University of Wisconsin-Madison on February 23, 2007.

All participants completed an anonymous demographic questionnaire to report their gender, age, racial/ethnic characteristics, educational level, and work experience. Participants also self-rated their biomedical understanding on a scale from 1 ("I rarely read texts on biomedical topics") through 4 ("I read and understand general medical articles") and their knowledge about diabetes mellitus on a scale from 1 ("very little") to 5 ("a good deal").

Table 2 (see the Results section) shows the characteristics of the sample obtained from this questionnaire.

Document Types

Clinical Trial

The first document type (see Textbox 1) was a description of a clinical trial entitled "Non invasive assessment of liver glycogen kinetics and ATP synthesis in type 1 diabetics", adapted from ClinicalTrials.gov (database trial identification number NCT00481598), the largest existing registry of clinical trials, maintained by the National Library of Medicine. This trial was

XSL•FO
RenderX

selected because it concerned diabetes mellitus, a common diagnosis, and because the documentation of the trial included a description of the study's purpose. This made it ideal for a study assessing consumer comprehension of text, as opposed to other dimensions of health literacy, such as understanding of tabular data, or numeracy in general. In fact, McCray and Ide [30] wrote early in ClinicalTrials.gov's history that one motivation for creating this website was the desire to make clinical trial information "available to individuals with serious or life-threatening diseases and conditions, *to other members of the public* [our emphasis], to health care providers, and to researchers" and available "in a form that can be readily understood by members of the public." Leroy and colleagues similarly chose a clinical trial document for a readability study because it represents "the most difficult language...that consumers will encounter and are expected to understand, that is, [a document] meant for them." [31]

Visit Notes

The second document type was a sample cardiology office Visit Notes document (Textbox 2) obtained from an online collection of sample medical transcripts at MedicalTranscriptionSamples.com. The site is a reference resource for medical transcriptionist training. The Visit Notes document was selected because of its focus on heart disease, a common consumer health concern. The document included the following sections: (1) history of present illness, (2) physical examination, (3) medications, (4) diagnoses, and (5) plan. A nurse practitioner reviewed the document and found it representative of office Visit Notes.

Textbox 1. Clinical Trial Document

NCT00481598 Non Invasive Assessment of Liver Glycogen Kinetics in Type1 Diabetics

Patients with Type 1 diabetes suffer from impaired postprandial hepatic glycogen storage and breakdown, if they are under poor glycaemic control. Poor glycogen storage in the liver puts these patients at risk of fasting hypoglycemia. Amelioration of glycaemic control could improve these abnormalities and thereby reduce the risk of hypoglycemia in these patients. The "gold standard" technique for the assessment of hepatic glycogen metabolism in humans, 13 C magnetic resonance spectroscopy (13C-MRS), is expensive and limited to a few centers worldwide. Aim 1 of our project is to establish a new assessment method for glycogen metabolism. This new method is based on oral administration of 2H2O and acetaminophen.

XSL•FO
RenderX

Textbox 2. Visit Notes document

History of Present Illness:

This 66-year-old white male was seen in my office on Month DD, YYYY. Patient was recently discharged from Doctors Hospital at Parkway after he was treated for pneumonia. Patient continues to have severe orthopnea, paroxysmal nocturnal dyspnea, cough with greenish expectoration. His exercise tolerance is about two to three yards for shortness of breath. The patient stopped taking Coumadin for reasons not very clear to him. He was documented to have recent atrial fibrillation. Patient has longstanding history of ischemic heart disease, end-stage LV systolic dysfunction, and is status post ICD implantation. Fasting blood sugar this morning is 130.

Physical Examination:

VITAL SIGNS: Blood pressure is 120/60. Respirations 18 per minute. Heart rate 75-85 beats per minute, irregular. Weight 207 pounds.

HEENT: Head normocephalic. Eyes, no evidence of anemia or jaundice. Oral hygiene is good.

NECK: Supple. JVP is flat. Carotid upstroke is good.

LUNGS: Severe inspiratory and expiratory wheezing heard throughout the lung fields. Fine crepitations heard at the base of the lungs on both sides.

CARDIOVASCULAR: PMI felt in fifth left intercostal space 0.5-inch lateral to midclavicular line. First and second heart sounds are normal in character. There is a II/VI systolic murmur best heard at the apex.

ABDOMEN: Soft. There is no hepatosplenomegaly.

EXTREMITIES: Patient has 1+ pedal edema.

Medications:

1. Ambien 10 mg at bedtime p.r.n.

2. Coumadin 7.5 mg daily.

3. Diovan 320 mg daily.

4. Lantus insulin 50 units in the morning.

5. Lasix 80 mg daily.

6. Novolin R p.r.n.

7. Toprol XL 100 mg daily.

8. Flovent 100 mcg twice a day.

Diagnosis:

1. Atherosclerotic coronary vascular disease with old myocardial infarction.

2. Moderate to severe LV systolic dysfunction.

3. Diabetes mellitus.

4. Diabetic nephropathy and renal failure.

5. Status post ICD implantation.

6. New onset of atrial fibrillation.

7. Chronic Coumadin therapy.

Plan:

1. Continue present therapy.

2. Patient will be seen again in my office in four weeks.

Study Conditions

We transformed the original documents three times to create the study conditions: health dictionary support (for the Dictionary condition), contextualized vocabulary support (for the Vocabulary condition), and coherence enhancement (for the Coherent condition). Each transformation is described below. For a summary comparison of characteristics of the original and transformed texts, see Table 1 in the Methods section, below.

Health Dictionary Support Transformation (Dictionary Condition)

We applied the predictive health term difficulty algorithm created by Zeng et al [5] to each document in order to identify terms unlikely to be familiar to consumers. Additionally, three nonclinician researchers independently extracted all potentially difficult health-related terms and expressions from the texts, adding them to the list of terms needing additional explanation. Finally, a nurse practitioner identified any remaining terms that were potentially problematic. These terms were selected for dictionary support.

XSL•FO
RenderX

"Difficult" terms were highlighted in blue in the text of the Dictionary condition. Terms in this condition had definitions provided in pop-up balloons activated by mousing over the text presented on a computer screen (see Figure 1). We culled definitions of terms from readily available Internet dictionary sources, such as Merriam-Webster's Medical Dictionary and others identified by Google's "define" function.

Contextualized Vocabulary Support Transformation (Vocabulary Condition)

This was similar to the transformation undertaken for the Dictionary condition described above, but in the Vocabulary condition, term definitions appearing in the pop-up balloons were edited by the nurse practitioner to specifically apply to the terms' contextual usage in the documents (see Figure 2).

Figure 1. Text selection with example of dictionary support.

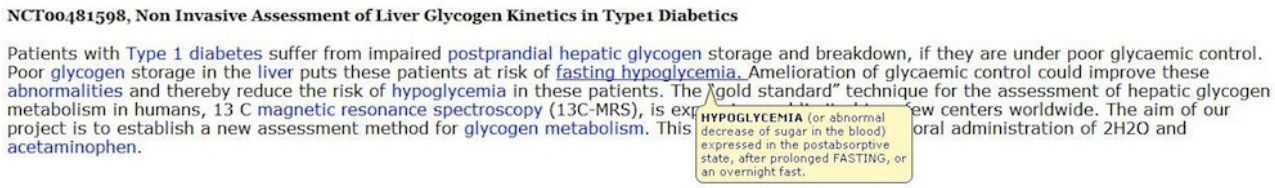
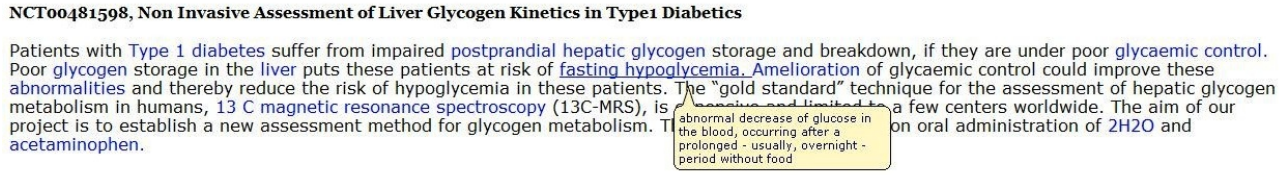


Figure 2. Text example with contextualized vocabulary support.



Coherence Enhancement (Coherent Condition)

This condition was developed in collaboration with the nurse practitioner, based on the principles outlined in the literature review above. We attempted to increase document coherence both at the local level (that is, between adjacent sentences) and at the global level (that is, across all sentences of the document), without altering the texts' graded readability level as measured by the Flesch-Kincaid formula [32]. Given the very different structures and the different intended original audiences for the two documents, we used two different procedures for improving their coherence.

While we attempted to target both local and global coherence, local coherence was not applicable to the sections of the Visit Notes that contained numbered items rather than sentences. This is because local coherence deals with sentence overlaps, often mentioning a concept from the previous sentence at the beginning of a new one; this makes it clear and unambiguous what various pronouns refer to (eg, does "it" refer to the heart or the procedure performed on it?). When text consists largely of bulleted or numbered lists, it is hard to do this kind of local coherence correction. For example, in a medications list made up of numbered sentence fragments, concepts mentioned in new sentences cannot be clearly linked to earlier sentences.

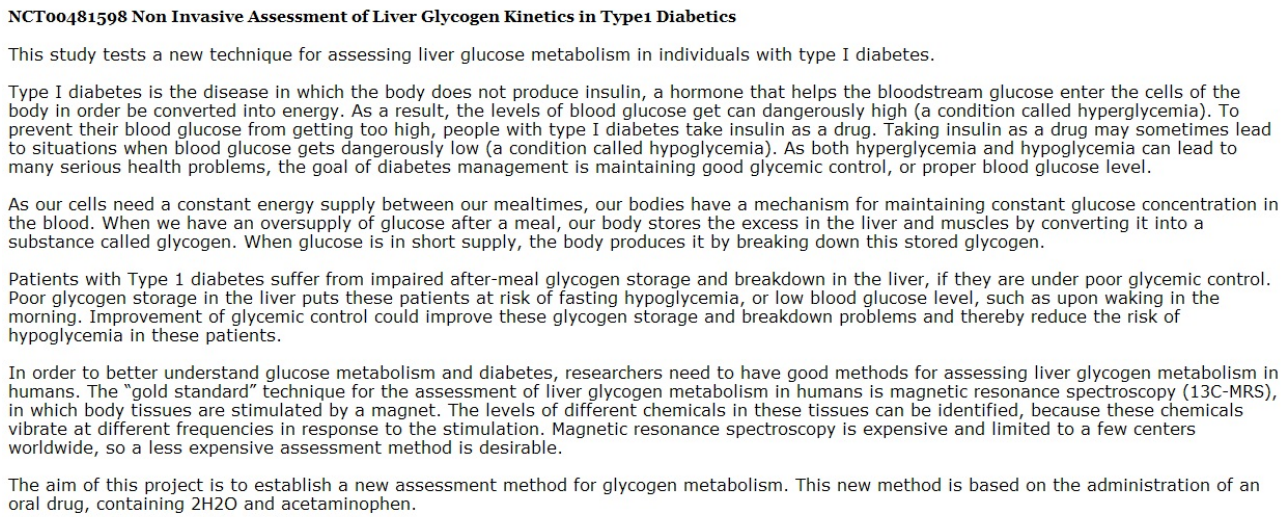
Global coherence, conversely, presents a contextual issue rather than a compositional one. We felt that coherence gaps in both documents had to do with the lay readers' insufficient background knowledge, leading to difficulty making inferences. In addition, potential coherence-related difficulties with the Visit Notes could be related to the topical organization and section and subsection headers in the document, a structure highly conventional and likely very familiar to medical professional authors and readers, but not to laypeople. The procedures by which we improved the coherence of the texts are described in detail below.

Clinical Trial Document Type

We first segmented this text into units of analysis, usually complete sentences. In some cases, complex sentences were divided into propositions, keeping intact phrases beginning with words such as "therefore" or "because." Next, we identified coherence gaps, defined as places where an inference was needed to comprehend each sentence on the basis of preceding sentences. Information was then added to the text, either by supplementing existing sentences or by adding new sentences that contained contextualized explanations. Examples of such added information include a missing background concept—for example, an explanation of the dangers of hypoglycemia—or the rationale behind the assessment procedure—for example, explaining the need to have good methods for measuring liver glycogen metabolism. Additionally, to make the clinical trial's research objectives more obvious, information about the purpose of the trial was rearranged from its original location so that it appeared in the opening sentence of the transformed document. Finally, to ensure local coherence, we checked the final text to ensure that the referents of pronouns were explicit. The coherence-transformed clinical trial text appears in Figure 3.

XSL•FO
RenderX

Figure 3. Clinical trial document with coherence enhancement.



NCT00481598 Non Invasive Assessment of Liver Glycogen Kinetics in Type1 Diabetics

This study tests a new technique for assessing liver glucose metabolism in individuals with type I diabetes.

Type I diabetes is the disease in which the body does not produce insulin, a hormone that helps the bloodstream glucose enter the cells of the body in order be converted into energy. As a result, the levels of blood glucose get can dangerously high (a condition called hyperglycemia). To prevent their blood glucose from getting too high, people with type I diabetes take insulin as a drug. Taking insulin as a drug may sometimes lead to situations when blood glucose gets dangerously low (a condition called hypoglycemia). As both hyperglycemia and hypoglycemia can lead to many serious health problems, the goal of diabetes management is maintaining good glycemic control, or proper blood glucose level.

As our cells need a constant energy supply between our mealtimes, our bodies have a mechanism for maintaining constant glucose concentration in the blood. When we have an oversupply of glucose after a meal, our body stores the excess in the liver and muscles by converting it into a substance called glycogen. When glucose is in short supply, the body produces it by breaking down this stored glycogen.

Patients with Type 1 diabetes suffer from impaired after-meal glycogen storage and breakdown in the liver, if they are under poor glycemic control. Poor glycogen storage in the liver puts these patients at risk of fasting hypoglycemia, or low blood glucose level, such as upon waking in the morning. Improvement of glycemic control could improve these glycogen storage and breakdown problems and thereby reduce the risk of hypoglycemia in these patients.

In order to better understand glucose metabolism and diabetes, researchers need to have good methods for assessing liver glycogen metabolism in humans. The "gold standard" technique for the assessment of liver glycogen metabolism in humans is magnetic resonance spectroscopy (13C-MRS), in which body tissues are stimulated by a magnet. The levels of different chemicals in these tissues can be identified, because these chemicals vibrate at different frequencies in response to the stimulation. Magnetic resonance spectroscopy is expensive and limited to a few centers worldwide, so a less expensive assessment method is desirable.

The aim of this project is to establish a new assessment method for glycogen metabolism. This new method is based on the administration of an oral drug, containing 2H2O and acetaminophen.

Visit Notes Document

The revision of this document involved a macrostructure analysis, performed by the nurse practitioner. This involved analyzing the relationship between sections of the document and the logic of the thematic organization of information within these sections. The nurse practitioner noted that grouping of diagnoses and complaints in the History of Present Illness and Diagnosis sections of the original document lacked a particular order. These complaints were accordingly regrouped into *heart-related*, *diabetes-related*, and related to *breathing difficulties*. "Chief complaint" was added to the breathing difficulties subheading. Just as the research objectives of the clinical trial were made more prominent in the transformed text, so for the Visit Note, medical concepts were explained in the body of the document. For example, in the Physical Examination section, test results were explained and interpreted (eg, by placement within or outside the normal range). In the Medications section, medications were regrouped by function; specific functions (eg, breathing problems; heart function and blood pressure) were explained, as were methods of action of individual medications.

Once we completed the coherence editing, we analyzed the text for readability level according to the Flesch-Kincaid Grade Level formula [32]. Based on these findings, adjustments were made to the Coherent condition of the clinical trial document, to ensure that its readability level remained comparable with that of the original text. In the case of the Visit Notes, the readability score for the coherent version was significantly higher than that of the original (see Table 1). As was noted early by Chapman et al [33], calculation of reading level using Flesch-Kincaid can be problematic for clinical text, because this formula relies partly on sentence length to establish difficulty, and medical documentation can be written in very short sentences. In our Visit Note, many sections of the original text were not written in complete sentences, resulting, in our estimation, in a deceptively low grade. In this case, we decided not to attempt matching readability levels, deeming that the Flesch-Kinkaid formula underestimated the difficulty of the original.

XSL•FO
RenderX

Table 1. Text characteristics of documents

Document type and condition	Number of words	Number of vocabulary definitions	Number of sentences	Grade level (Flesch-Kincaid)
Clinical trial				
Original	108	NA[a]	6	14.1
Dictionary	[b]	12	[b]	[b]
Vocabulary	[b]	12	[b]	
Coherent	394	NA[a]	18	13.8
Visit Notes				
Original	326		43	9.5
Dictionary	[b]		[b]	
Vocabulary	[b]	12	[b]	
Coherent	1219		66	11.3

[a] Not applicable.

[b] Dictionary- and Vocabulary-enhanced versions had the same number of words and same Flesch-Kincaid Grade Readability Level and sentences as the original versions.

Procedure

Study participants worked on individual computers; a research assistant served as proctor, observing at all times to ensure that work was done individually. Participants were randomly assigned to one of the four conditions (Original, Dictionary, Vocabulary, or Coherent). It was explained to all participants that some of them would see balloon features in their documents, and that they should feel free to take advantage of these features.

The order of presentation of the clinical trial and Visit Notes documents was randomized among participants. After completing the anonymous demographic questionnaire, participants read their first document on the computer screen. After a waiting period of 10 minutes, participants wrote their recollection of the text they read in this document using Microsoft Word. They were instructed to retell the document they had just read as if they were making the information available to a person who had never seen it before. In addition, participants in the second, but not the first, cohort answered an open-ended questionnaire about the text they read (see Textbox 3). This procedure was then repeated for each participant's second document.

Observation during the session indicated that participants did indeed invoke the balloon features. Participants were allowed to take as long as they required to "retell" each text; the modal time to completion was 20 minutes in both cohorts. The time period was selected to be sufficient for all participants to complete the task without pressure, regardless of the length of the stimulus text. All participants were able to finish their work before the time elapsed.

Textbox 3. Open-Ended Questionnaire for Clinical Trial Text.

1. Who is being recruited for the study described in this paragraph?
2. This paragraph mentions measuring something. What is the thing that is being measured?
3. Why is it important to measure this thing?
4. Many health problems are associated with diabetes. Which particular health problem is the main focus of this text?
5. What is the innovation of the research described in this text?

Coding and Statistical Analysis

There were three outcomes of interest: (1) number of propositions recalled, (2) open-ended questionnaire score, and (3) number of errors made by participants. These were collected over the four study conditions for each of the two document types. We assessed the effect of the conditions on the outcomes separately for each document type.

Demographic Questionnaire Analysis

Demographic variables were summarized by frequency and percentage or median and interquartile range (IQR) based on the distribution type of each variable. We compared demographic factor variables between the four groups with Fisher exact tests. We compared demographic score variables between the four groups with Kruskal-Wallis tests. The Kruskal-Wallis test was used because it is the nonparametric test for comparing more than two groups. All demographic comparisons were insignificant so no pairwise comparisons were made.

Propositional Analysis

We followed the standard procedure of segmenting original versions of each text into propositions, or basic units of analysis

XSL•FO
RenderX

corresponding to two concepts connected by a relationship (eg, [antigen] attacks [immune system]) or a concept with a modifier (eg, severe [pain]) [34]. Disagreements about whether a particular statement constituted a proposition were resolved via discussions among three of the authors (AK, CAS, and PWD). Scoring was based on participants' recall of the propositions of the original texts. Lists of propositions found in the original texts were used as scoring sheets against which to analyze participants' recall.

Each transcript was scored to indicate the presence or absence of the original text's proposition in the retelling. The coding guide was developed through discussions, using a pilot (training) dataset. We obtained the pilot retellings from the our colleagues and family members with demographics similar to the participants'. Two raters (AK and CAS) scored three randomly selected pilot retellings of each document. The analysis of interrater reliability yielded kappa coefficients of .73 (substantial agreement), .8 (almost perfect agreement), and .83 (almost perfect agreement) for the Visit Notes and .71 (substantial agreement), .76 (substantial agreement), and .8 (almost perfect agreement) for the clinical trial. Disagreements were resolved via discussions, following which AK and CAS each scored half of the protocols. The transcripts were scored in random order and the scorer was blind to the condition being scored [35].

Open-Ended Questionnaire

We administered an open-ended questionnaire to each participant in Cohort 2, one questionnaire for each document type for a total of two questionnaires per participant. (For an example of the clinical trial document's open-ended questionnaire, see Textbox 3 above). Authors CAS and AK jointly coded all the questionnaires and resolved disagreements through mutual discussion. The clinical trial questionnaire was scored by assigning each answer a score of 0, 1, or 2, reflecting the accuracy and completeness of participants' answers; for the Visit Notes questionnaire, since answers reflected retention and understanding of much more granular information, a point was awarded for each medication, diagnosis, etc. recalled correctly by the participant.

Statistical Analysis

Similar statistical analysis was performed for (1) the number of propositions recalled, (2) open-ended questionnaire score, and (3) number of errors. For each variable, initially, a Kruskal-Wallis test was performed to test for differences in the outcomes based on the four study conditions. In those circumstances in which the Kruskal-Wallis test was insignificant, the initial hypothesis that there would be no difference between the Original and Dictionary conditions was tested with a Wilcoxon rank sum test. If this test was also insignificant, then these two groups were combined and Kruskal-Wallis analysis was rerun comparing the three condition groups as follows: (1) Original + Dictionary, (2) Vocabulary, and (3) Coherent. If this Kruskal-Wallis test was significant, then pairwise Wilcoxon rank sum tests were conducted with Holm adjusted *P* values for multiple comparisons. All comparisons were conducted at an alpha level of .05.

In addition, post hoc power analysis was done for comparison of the Visit Notes total open-ended questionnaire scores among the four conditions. This analysis was done for the Visit Notes, but not for the clinical trial, because for the Visit Notes, the distribution of the medians for the four conditions showed a steady trend in the expected direction.

Results

Table 2 shows results of the experiment and characteristics of the participants

XSL•FO
RenderX

Table 2. Characteristics of participants

Variable	Intervention group				*P* value
	Original	Dictionary	Vocabulary	Coherent	
Gender, n (%)					
Female	17 (85)	15 (75)	17 (85)	15 (75)	.59
Male	3 (15)	5 (25)	3 (15)	5 (25)	
Age (years), n (%)					.48
<30	15 (79)	14 (70)	12 (60)	15 (75)	
30–39	2 (11)	4 (20)	4 (20)	0 (0)	
40–49	1 (5)	1 (5)	2 (10)	4 (40)	
50–65	1 (5)	1 (5)	1 (5)	1 (5)	
>65	0 (0)	0 (0)	1 (5)	0 (0)	
Education level attained, n (%)					.91
High school	3 (15)	2 (10)	3 (15)	2 (10)	
College degree	12 (60)	12 (60)	10 (50)	12 (60)	
Master's	5 (25)	5 (25)	7 (35)	4 (20)	
>Master's	0 (0)	1 (5)	0 (0)	2 (10)	
Degree type a , n (%)					1.00
Health-related	1 (5)	1 (6)	1 (6)	1 (5)	
Nonhealth-related	18 (95)	17 (94)	17 (94)	19 (95)	
Biomedical knowledge					
Median (IQR)[b]	1.5 (1.0–3.0)	2.0 (1.0–2.0)	2.0 (1.0–2.0)	2.0 (2.0–3.0)	.15
Diabetic knowledge					
Median (IQR)	2.0 (1.0–3.0)	2.0 (1.8–3.0)	2.0 (2.0–3.3)	3.0 (2.0–3.3)	.72

[a] Of highest earned degree.

[a] Interquartile range.

Number of Original Text Propositions Recalled

The effect of the version on the number of the original text propositions recalled was assessed separately for each document type, clinical trial and Visit Notes alike. Both document types showed insignificant differences between the Original and Dictionary conditions ($P = .65$, $P = .48$, respectively). The two conditions were combined for the subsequent analysis.

For the clinical trial text, the effect of the condition was not significant in any of the comparisons, suggesting no differences in recall, despite the varying levels of support ($P = .84$). For the Visit Note, however, we found a significant difference in the median total propositions recalled between the Coherent and the (Original + Dictionary) conditions ($P = .04$). This suggests that participants in the Coherent condition recalled more of the original Visit Notes content than did participants in the Original and the Dictionary conditions combined. No comparisons involving the Vocabulary condition were significant. Median, IQR, and range for the number of propositions recalled for each document type are presented in Table 3.

XSL•FO
RenderX

Table 3. Total propositions recalled

Document type and condition	n	Median	IQR[a]	Range	Contrast	*P* value
Clinical trial						
Original	20	8.5	6.75–12.25	4–21	O[b] vs D[c]	.63
Dictionary	20	9.0	7.0–13.25	5–23		
Combined (O[b] + D[c])	40	9.0	7.0–13.0	4–23	Kruskal-Wallis	.84
Vocabulary	20	10.0	6.75–12.25	2–18		
Coherent	20	10.5	7.75–13.5	4–18		
Visit Notes						
Original	20	17.5	14.0–21.25	9–39	O[b] vs D[c]	.48
Dictionary	20	20.0	16.5–23.25	4–41		
Combined (O[b] +D[c])	40	19.0	15.0–22.0	4–41	O[b] + D[c] vs V[d]	.36[e]
Vocabulary	20	22.5	15.75–32.75	5–50	V[d] vs C[f]	.62[e]
Coherent	20	25.5	20.5–33.25	13–41	O[b] + D[c] vs C[f]	.04[e]

[a] Interquartile range.

[b] Original.

[c] Dictionary.

[d] Vocabulary.

[e] Holm adjusted *P* values for multiple comparisons.

[f] Coherent.

Open-Ended Questionnaire Scores

This comparison involved the effect of the conditions on the open-ended questionnaire scores. For both text types, the initial Kruskal-Wallis comparison of the Original and Dictionary conditions was insignificant (clinical trial: $P = .70$, Visit Note: $P = .36$), so the two conditions were combined. The analysis found no significant effect of the text version in any of the clinical trial comparisons ($P = .86$). The effect of the text version for the Visit Notes also did not reach significance ($P = .20$). Median, IQR, and range for the number of main ideas for each document type are presented in Table 4.

Table 4. Open-ended questionnaire scores

Document type and condition	N	Median	IQR[a]	Range	Contrast	*P* value
Clinical trial						
Original	20	6.0	3.25–8.0	1–10	O[b] vs D[c]	.70
Dictionary	20	6.0	4.5–8.0	3–10		
Combined (O[b] + D[c])	40	6.0	3.75–8.0	1–10	Kruskal-Wallis	.86
Vocabulary	20	5.5	5.0–7.0	3–8		
Coherent	20	6.0	4.25–6.0	3–9		
Visit Notes						
Original	20	11	10.25–14.25	7–20	O[b] vs D[c]	.36
Dictionary	20	13.5	12.25–15.5	9–16		
Combined (O[b] + D[c])	40	12.5	10.75–15.25	7–20	Kruskal-Wallis	.20
Vocabulary	20	14.0	13.0–18.75	7–20		
Coherent	20	15.0	14.25–15.0	10-18		

[a] Interquartile range.

[b] Original.

[c] Dictionary.

XSL•FO
RenderX

Because the open-ended questionnaire was added to the study after the first half of the participants completed the study, the data sample was small, consisting of 10 participants per condition. We performed a post hoc power analysis for the test of differences between the four treatment conditions for Visit Notes open-ended questionnaire scores. With the assumption of normal data with means 11, 13.5, 14, and 15, which were the median values seen in the actual data, and overall standard deviation of 3.4, the post hoc power analysis indicated that we had only 57% power to find a difference with 10 participants per condition. To achieve adequate 80% power to detect a difference, under the normality assumption, we would have needed 16 participants per condition. Even though the data are nonnormal, they are only slightly skewed from normality, and this would only minimally increase the needed sample size for sufficient power. Median, IQR, and range for open-ended questionnaire scores for each document type are presented in Table 4.

Number of Errors

The initial Kruskal-Wallis comparison of the Original and Dictionary conditions was insignificant (clinical trial: $P = .20$, Visit Notes: $P = .91$), so the two conditions were combined. The analysis found no significant differences, regardless of the document type. Median, IQR, and range for the number of errors for each document type are presented in Table 5.

Table 5. Total errors made

Document type and condition	N	Median	IQR[a]	Range	Contrast	*P* value
Clinical trial						
Original	10	1.5	0.75–3.0	0–5	O^b vs D^c	.99
Dictionary	10	1.0	1.0–3.0	0–6		
Combined (O^b + D^c)	20	1.0	1.0–3.0	0–6	Kruskal-Wallis	.47
Vocabulary	10	2.0	1.0–2.0	0–3		
Coherent	10	2.0	1.0–3.25	0–6		
Visit Notes						
Original	10	2.5	1.0–3.25	0–7	O^b vs D^c	.91
Dictionary	10	2.0	2.0–4.0	0–4		
Combined (O^b + D^c)	20	2.0	1.0–4.0	0–7	Kruskal-Wallis	.25
Vocabulary	10	2.0	1.0–3.25	0–5		
Coherent	10	3.5	2.0–5.0	0–10		

[a] Interquartile range.

Discussion

The results of this study expand our understanding of consumer difficulties with the technical language of medicine. Much research in this area has focused on terminology bridge solutions through technologies such as the Unified Medical Language System. Slaughter et al [36] looked at consonance of patient symptom expressions with nurses' terminology in the medical record, but the goal of this research was to understand differences, not to measure incomprehension. Similarly, Hong et al [37] compared terminology in an electronic health record system with patient-friendly terms in the same system to find consonance between the two.

On a purely lexical basis, a translation from clinical to consumer language is appealing. Unfortunately, making complex clinical concepts clearer to laypeople requires more than a dictionary. The physician's lack of time to explain concepts found in medical records was an often-cited criticism in the early literature concerning patient access to those records [8,11]. For this reason, two early studies built time and personnel resources into their design to avoid this problem. Golodetz et al [38] explained "necessary technical language" to the 60% of their study participants requesting this assistance (N = 103). Stein et al [39] provided their psychiatric patient participants with at least one nursing staff member to help explain terminology. Fischbach et al [40] surveyed the depth of the problem by designing a study in which patients and providers collaborated on authorship of the medical record: 20 patients with mixed diagnoses were asked to initiate and formulate their own problem list, with four providers suggesting modifications; both parties then worked together to write continuation notes (symptoms, clinical findings, and assessments). Fischbach et al found that physicians' prospective worries about the time required to effectively communicate were entirely justified; these coauthoring consultations took as much as 50% longer than traditional visits; but these researchers saw value in incorporating the patient perspective into the health care documentation process. Participation in the coauthorship process "helped to eliminate serious misconceptions on the part of the patients." In fact, a new language of cooperation was described as emerging out of this dialogue: "[T]he requirement for collaborative writing, which necessitated constant negotiation and feedback, created a meld of medical jargon and layman's slang into a mutually useful language" (p 3). [40].

What do the results of this study tell us? We present here our original hypotheses:

XSL•FO
RenderX

H1: Readers' comprehension of a text enhanced by providing standard, off-the-shelf dictionary definitions will not be significantly greater than their comprehension of the original text.

As expected, there was no difference in comprehension (as measured by the recall, answers to open-ended questions, or the number of errors) between the Original and Dictionary conditions. Comprehension was measured by the participants' recall, their answers to open-ended questions, and the number of errors they made.

This supports the contention of Zeng and Tse [16] that the simple provision of a dictionary does not improve reader comprehension. However, it is important to remember that the dictionary is only a vehicle by which vocabulary is transported; vocabulary is the real problem, not the dictionary itself. Dictionary definitions may indeed be simple, clear, and likely to help the user; for the medical words we reviewed, however, the typical consumer dictionary was found to be extremely unuseful. For example, the National Library of Medicine's consumer health website, MedlinePlus, is a portal intended explicitly for laypeople and not for health care professionals or researchers; among its licensed resources is the Merriam-Webster Medical Dictionary. This dictionary's definition for acetaminophen reads as follows:

a crystalline compound C8H9NO2that is a hydroxy derivative of acetanilide and is used in chemical synthesis and in medicine instead of aspirin to relieve pain and fever—called also paracetamol*; see liquiprin, panadol, tylenol* [41]

Readers who does not know what acetaminophen is are unlikely to be assisted by this information.

H2: Readers' comprehension of a vocabulary-enhanced text will be significantly greater than in the Original and Dictionary conditions combined.

This hypothesis was not supported. A specifically contextualized vocabulary developed for the purposes of this text did not improve comprehension, as assessed by any of the three comprehension measures. The lack of positive effect of a carefully constructed, clear vocabulary is counterintuitive. This result may be attributable to our choice of very complex medical texts for the study. The conceptual density of these texts may have created coherence gaps that were too large to be ameliorated by vocabulary definitions.

H3: Readers' comprehension of a text with improved coherence will be significantly greater than in the Original plus Dictionary condition.

This hypothesis is partially supported, for the Visit Notes document but not for the clinical trial document. The *P* value for the Visit Notes in the Coherent condition compared with the Original plus Dictionary conditions is significant at .04. For this particular hypothesis, then, the researcher's glass is half empty *and* half full. Many cognitive studies in other fields have shown that coherence is a factor affecting comprehension. Our results show that this is true for the Visit Notes document, a particularly impressive finding because, as discussed above, improving coherence of this document required making an already long text even longer—while the original Visit Notes document was 326 words long, the version with enhanced coherence totaled 1219 words. Despite this fourfold increase in length, the more coherent document still managed to hold the participants' attention. Examples showing the difference between a participant with high recall in the Coherent condition and one with a low recall with the Original text condition appear in Textboxes 4 and 5. Each example is a description of the cardiac problems remembered from the Visit Notes.

For the clinical trial document, however, this is still not a promising result. While the median recall was increased from 9.0 to 10.5 propositions in the Coherent condition over the Original and Dictionary conditions, the error rate remained the same in the Coherent condition as in the other conditions; that is, no matter what was done to the text, the number of errors remained constant. The clinical trial document, then, was apparently simply so difficult, and so short, that nothing was able to make it easier to read.

H4: Readers' comprehension of the Coherent condition will be significantly greater than in the Vocabulary condition.

This hypothesis was rendered irrelevant by the overall lack of significant comprehension improvement in the Vocabulary condition. Our expectation had been that both the Vocabulary and the Coherent conditions would improve comprehension compared with the Original and Dictionary conditions, with the gain being greater for the Coherent conditions. In this study, however, the improvement was observed only for the Coherent condition (and then only for the Visit Notes text).

Textbox 4. Excerpt of Visit Notes Text About Cardiac Problems Composed by Participant 28 Showing High Recall of Propositions [total of 43] in the Coherent Condition

Heart:

1. The blood vessels are tightening as the result of a build up of cholesterol.

2. The patients heart beats irregularly

3. The patient has a pacemaker device to help control the hearts beat, this works by sending an electric pulse when the patients heart gets off beat.

4. There is a particular weakness in the left ventricle of the patient's heart.

5. The patient is on blood thinners to reduce the risks of clotting which are a special threat for patients having suffered a heart attack, such as this patient.

XSL•FO
RenderX

Textbox 5. Excerpt of Visit Notes Text About Cardiac Problems Composed by Participant 5 Showing Very Low Recall of Propositions [total of 16] in the Original Text Condition

Patient Visit
b) Irregular heartbeat, wheezing, strong carotid pulse, soft abdomen, good oral hygiene, heart murmur, supple neck

Limitations

This study has limitations that may restrict the generalizability of its results. These include its small size (80 participants) and the educational background of the research participants: 90% were college graduates. This educational background, however, does allow us to make the suggestion that people with less education could have performed even more poorly. Additionally, our conclusions may be confounded by the fact that we tested only one clinical trial and only one Visit Notes document. It is difficult to say, for example, whether a clinical trial involving rheumatoid arthritis and Visit Notes involving pregnancy would have evoked different readability responses in our participants.

Conclusion and Future Directions

We conclude by reviewing the findings of this study and examining their implications for future work. The practical significance of this study lies in showing the full extent of the difficulty and labor intensiveness of improving comprehension of clinical documents. This work explores cognitive characteristics of the reader–text match that show why commonly attempted solutions—lowering readability scores and providing dictionary definitions—are not sufficient. It also points to strategies for intervention that merit future research attention. Much research effort could be directed at (1) identifying aspects of coherence that are particularly relevant for comprehending complex medical texts, and (2) seeking automatic tools that can aid in document revision. Statistically, we have shown that improving coherence of typical clinical documents has a small effect on consumer comprehension, but this task is not scalable with automated solutions and would be impractical with manual solutions. Perhaps a promise of automation scalability lies in an iterative hybrid approach, where automated textual analysis for coherence is followed by manual editing, which is then rechecked with an automated tool. While automatic text editing is still a matter of the distance future, validated automated tools capable of distinguishing between high- and low-coherence versions of textual documents do exist [26]. Unfortunately, while an automated approach is well fitted for analyzing indices of local coherence, such as argument overlap, it is not capable of assessing many aspects of global coherence, such as the appropriateness of topic sentences and the background information level. In the case of knowledge-rich texts, such as medical documents, increasing local coherence alone is likely to be insufficient. Further research is needed using texts from more diverse clinical domains and more heterogeneous participants, including actual patients.

Second, it is interesting that the coherence-enhanced Visit Notes document was able to hold readers' attention despite the fact that increasing coherence almost quadrupled the size of the document. This finding has implications not only for coherence, but also for text construction itself. It may be the narrative format that allows lay readers to form a more coherent story. Thinking of the medical record as narrative is a well-established trope in the medical humanities; Epstein, for example, writing about the development of genetics, points to the importance of the physician as writer: "a chronicler of bodily events and systematic narrator of particular phenomena in a particular context" [42]. Kennedy points to the "case...as the predominant form of medical narrative" and argues that it cannot be understood "aside from its involvement with literary discourse" [43] Recent work on illness narratives constructed from diaries—written by both nurses [44] and patients [45]—reveals that narrative structure assists participants in health care in sense-making—constructing a coherent account of the illness. In fact, considered in this light, the 30-year-old study by Fischbach et al [40] may have been as much about narrative as it was about medical record co-construction.

Finally, our results suggest that given the difficulty of engineering comprehensibility of clinical text, the most useful informatics tools will be those that can support the physicians, nurses, and patient educators tasked with making clinical information understandable to patients. These health care professionals use a repetitive cycle of explaining concepts, asking questions to ensure that patients comprehend, and explaining again. If the attainment of coherence is the end result of an iterative process, no single instance of a static document will solve the coherence problem.

Acknowledgments

The authors gratefully acknowledge the helpful assistance of Hyeoneui Kim, RN, MPH, PhD, during the development of the various text conditions, and Qing Zeng-Treitler, PhD; funding support by NIDDK 7R01DK075837; and the Intramural Research Program of the National Library of Medicine.

Conflicts of Interest

None declared

References

XSL•FO
RenderX

1. McCray AT. Promoting health literacy. J Am Med Inform Assoc 2005;12(2):152-163 [FREE Full text] [doi: 10.1197/jamia.M1687] [Medline: 15561782]
2. Davis TC, Crouch MA, Wills G, Miller S, Abdehou DM. The gap between patient reading comprehension and the readability of patient education materials. J Fam Pract 1990 Nov;31(5):533-538. [Medline: 2230677]
3. Davis TC, Mayeaux EJ, Fredrickson D, Bocchini JA, Jackson RH, Murphy PW. Reading ability of parents compared with reading level of pediatric patient education materials. Pediatrics 1994 Mar;93(3):460-468. [Medline: 8115206]
4. Nielsen-Bohlman L, Panzer AM, Kindig DA, Committee on Health Literacy, Board on Neuroscience and Behavioral Health, Institute of Medicine. Health Literacy: A Prescription to End Confusion. Washington, DC: National Academies Press; 2004.
5. Zeng Q, Kim E, Crowell J, Tse T. A text corpora-based estimation of the familiarity of health terminology. In: Maojo V, Martin-Sanchez F, Pereira A, editors. Biological and Medical Data Analysis: 6th International Symposium, ISBMDA 2005, Aveiro, Portugal, November 10-11, 2005, Proceedings. Berlin: Springer; 2005:184-192.
6. Britten N, Bartholomew J, Morris R, Zander L. Consultants' and patients' views about patient access to their general practice records. J R Soc Med 1991 May;84(5):284-287. [Medline: 2041006]
7. Chamberlain G. Should patients read their own medical records? J R Soc Med 1987 Sep;80(9):541-542. [Medline: 3681865]
8. Fisher B, Britten N. Patient access to records: expectations of hospital doctors and experiences of cancer patients. Br J Gen Pract 1993 Feb;43(367):52-56. [Medline: 8466775]
9. Lovell A. In whose interest is access to medical records withheld? Br J Obstet Gynaecol 1987 Jul;94(7):609-611. [Medline: 3620410]
10. Short D. Some consequences of granting patients access to consultants' records. Lancet 1986 Jun 7;1(8493):1316-1318. [Medline: 2872441]
11. Westin AF. Medical records: should patients have access? Hastings Cent Rep 1977 Dec;7(6):23-28. [Medline: 924782]
12. Tomkins CS, Braid JJ, Williams HC. Do dermatology outpatients value a copy of the letter sent to their general practitioner? In what way and at what cost? Clin Exp Dermatol 2004 Jan;29(1):81-86. [Medline: 14723730]
13. Nijland N, van Gemert-Pijnen J, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]
14. Keselman A, Smith CA, Divita G, Kim H, Browne AC, Leroy G, et al. Consumer health concepts that do not map to the UMLS: where do they fit? J Am Med Inform Assoc 2008 Aug;15(4):496-505 [FREE Full text] [doi: 10.1197/jamia.M2599] [Medline: 18436906]
15. Smith CA. Consumer language, patient language, and thesauri: a review of the literature. J Med Libr Assoc 2011 Apr;99(2):135-144. [doi: 10.3163/1536-5050.99.2.005] [Medline: 21464851]
16. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. J Am Med Inform Assoc 2006 Feb;13(1):24-29 [FREE Full text] [doi: 10.1197/jamia.M1761] [Medline: 16221948]
17. Patrick TB, Monga HK, Sievert ME, Houston Hall J, Longo DR. Evaluation of controlled vocabulary resources for development of a consumer entry vocabulary for diabetes. J Med Internet Res 2001;3(3):E24 [FREE Full text] [doi: 10.2196/jmir.3.3.e24] [Medline: 11720966]
18. Leroy G, Miller T. Perils of providing visual health information overviews for consumers with low health literacy or high stress. J Am Med Inform Assoc 2010;17(2):220-223 [FREE Full text] [doi: 10.1136/jamia.2009.002717] [Medline: 20190068]
19. Velardi P, Navigli R, d'Amadio P. Mining the Web to create specialized glossaries. IEEE Intell Syst 2008;23(5):18-25. [doi: 10.1109/MIS.2008.88]
20. Diefenbach MA, Butz BP. A multimedia interactive education system for prostate cancer patients: development and preliminary evaluation. J Med Internet Res 2004 Jan 21;6(1):e3 [FREE Full text] [doi: 10.2196/jmir.6.1.e3] [Medline: 15111269]
21. McNamara DS. Reading both high-coherence and low-coherence texts: effects of text sequence and prior knowledge. Can J Exp Psychol 2001 Mar;55(1):51-62. [Medline: 11301728]
22. Albrecht JE, O'Brien EJ. Updating a mental model: maintaining both local and global coherence. J Exp Psychol Learn Mem Cogn 1993;19(5):1061-1070. [doi: 10.1037/0278-7393.19.5.1061]
23. Kintsch W. Comprehension: A Paradigm for Cognition. Cambridge, UK: Cambridge University Press; 1998.
24. McNamara DS, Kintsch E, Songer NB, Kintsch W. Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. Cogn Instr 1996;14(1):1-41. [doi: 10.1207/s1532690xci1401_1]
25. Britton BK, Gulgoz S. Using Kintsch's computational model to improve instructional text: effects of repairing inference calls on recall and cognitive structures. J Educ Psychol 1991;83(3):329-345. [doi: 10.1037/0022-0663.83.3.329]
26. McNamara DS, Louwerse MM, McCarthy PM, Graesser AC. Coh-Metrix: capturing linguistic features of cohesion. Discourse Process 2010;47(4):292-330. [doi: 10.1080/01638530902959943]
27. Vidal-Abarca E, Martínez G, Gilabert R. Two procedures to improve instructional text: effects on memory and learning. J Educ Psychol 2000;92(1):107-116. [doi: 10.1037/0022-0663.92.1.107]

XSL•FO
RenderX

28. Linderholm T, Everson MG, van den Broek P, Mischinski M, Crittenden A, Samuels J. Effects of causal text revisions on more- and less-skilled readers' comprehension of easy and difficult narrative texts. Cogn Instr 2000;18(4):525-556. [doi: 10.1207/S1532690XCI1804_4]
29. Kools M, Ruiter RA, van de Wiel MW, Kok G. Increasing readers' comprehension of health education brochures: a qualitative study into how professional writers make texts coherent. Health Educ Behav 2004 Dec;31(6):720-740. [doi: 10.1177/1090198104263340] [Medline: 15539544]
30. McCray AT, Ide NC. Design and implementation of a national clinical trials registry. J Am Med Inform Assoc 2000 Jun;7(3):313-323 [FREE Full text] [Medline: 10833169]
31. Leroy G, Helmreich S, Cowie JR, Miller T, Zheng W. Evaluating online health information: beyond readability formulas. AMIA Annu Symp Proc 2008:394-398. [Medline: 18998902]
32. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. Naval Technical Training Command. 1975 Feb. Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy Enlisted Personnel URL: http://digitalcollections.lib.ucf.edu/u/?/IST,26253 [accessed 2011-10-26] [WebCite Cache ID 62jmYwzsB]
33. Chapman WW, Fiszman M, Frederick PR, Chapman BE, Haug PJ. Quantifying the characteristics of unambiguous chest radiography reports in the context of pneumonia. Acad Radiol 2001 Jan;8(1):57-66. [doi: 10.1016/S1076-6332(03)80744-4] [Medline: 11201458]
34. Arocha JF, Wang D, Patel VL. Identifying reasoning strategies in medical decision making: a methodological guide. J Biomed Inform 2005 Apr;38(2):154-171. [doi: 10.1016/j.jbi.2005.02.001] [Medline: 15797004]
35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
36. Slaughter L, Ruland C, Rotegård AK. Mapping cancer patients' symptoms to UMLS concepts. AMIA Annu Symp Proc 2005:699-703. [Medline: 16779130]
37. Hong Y, Ehlers K, Gillis R, Patrick T, Zhang J. A usability study of patient-friendly terminology in an EMR system. Stud Health Technol Inform 2010;160(Pt 1):136-140. [Medline: 20841665]
38. Golodetz A, Ruess J, Milhous RL. The right to know: giving the patient his medical record. Arch Phys Med Rehabil 1976 Feb;57(2):78-81. [Medline: 1083223]
39. Stein EJ, Furedy RL, Simonton MJ, Neuffer CH. Patient access to medical records on a psychiatric inpatient unit. Am J Psychiatry 1979 Mar;136(3):327-329. [Medline: 420331]
40. Fischbach RL, Sionelo-Bayog A, Needle A, Delbanco TL. The patient and practitioner as co-authors of the medical record. Patient Couns Health Educ 1980;2(1):1-5. [Medline: 10294999]
41. Merriam-Webster Medical Dictionary. MedlinePlus. 2011. Acetaminophen URL: http://www.merriam-webster.com/medlineplus/acetaminophen [accessed 2011-10-28] [WebCite Cache ID 62m43o2au]
42. Epstein J. Altered Conditions: Diseases, Medicine, and Storytelling. New York, NY: Routledge; 1995.
43. Kennedy MA. A Curious Literature: Reading the Medical Case History From the Royal Society to Freud [dissertation]. Providence, RI: Brown University; 2000.
44. Egerod I, Bagger C. Patients' experiences of intensive care diaries--a focus group study. Intensive Crit Care Nurs 2010 Oct;26(5):278-287. [doi: 10.1016/j.iccn.2010.07.002] [Medline: 20692158]
45. Egerod I, Christensen D. A comparative study of ICU patient diaries vs. hospital charts. Qual Health Res 2010 Oct;20(10):1446-1456. [doi: 10.1177/1049732310373558] [Medline: 20562254]

Edited by G Eysenbach; submitted 06.05.11; peer-reviewed by D Kaufman, S Jamshed, A Lewis; comments to author 06.06.11; revised version received 08.07.11; accepted 10.08.11; published 02.12.11.

Please cite as:
Smith CA, Hetzel S, Dalrymple P, Keselman A
Beyond Readability: Investigating Coherence of Clinical Text for Consumers
J Med Internet Res 2011;13(4):e104
URL: http://www.jmir.org/2011/4/e104/
doi:10.2196/jmir.1842
PMID:22138127

©Catherine Arnott Smith, Scott Hetzel, Prudence Dalrymple, Alla Keselman. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 02.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Analysis of the Definition and Utility of Personal Health Records Using Q Methodology

Jeongeun Kim[1,2,3,4], RN, INS, PhD; David W Bates[4,5,6], MSc, MD

[1]College of Nursing, Seoul National University, Seoul, Republic Of Korea

[2]Research Institute of Nursing Science, Seoul National University, Seoul, Republic Of Korea

[3]Interdisciplinary Program of Medical Informatics, Seoul National University, Seoul, Republic Of Korea

[4]Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, United States

[5]Harvard Medical School, Boston, MA, United States

[6]Partners Information Systems, Partners HealthCare, Boston, MA, United States

Corresponding Author:
Jeongeun Kim, RN, INS, PhD
College of Nursing
Seoul National University
28 Yongon-dong Jongno-gu
Seoul, 110-799
Republic Of Korea
Phone: 82 2 740 8483
Fax: 82 2 765 4103
Email: kim0424@snu.ac.kr

Abstract

Background: Personal health records (PHRs) remain a relatively new technology and concept in practice even though they have been discussed in the literature for more than 50 years. There is no consensus on the definition of a PHR or PHR system even within the professional societies of health information technology.

Objective: Our objective was to analyze and classify the opinions of health information professionals regarding the definitions of the PHR.

Method: Q methodology was used to explore the concept of the PHR. A total of 50 Q-statements were selected and rated by 45 P-samples consisting of health information professionals. We analyzed the resulting data by using Q methodology-specific software and SPSS.

Result: We selected five types of health information professionals' opinions: type I, public interest centered; type II, health information standardization centered; type III, health consumer centered; type IV, health information security centered; and type V, health consumer convenience centered. The Q-statements with the highest levels of agreement were as follows: (1) the PHR is the lifetime record of personal health information, (2) the PHR is the representation of health 2.0, and (3) security is the most important requirement of the PHR. The most disagreed-with Q-statements were (1) the PHR is a paper-based system, and (2) it is most effective to carry the PHR information in USB storage.

Conclusion: Health information professionals agree that PHRs should be lifetime records, that they will be useful as more information is stored electronically, and that data security is paramount. To maximize the benefits of PHR, activation strategies should be developed and extended across disciplines and professionals so that patients begin to receive the benefits associate with using PHRs.

(J Med Internet Res 2011;13(4):e105) doi:10.2196/jmir.1781

KEYWORDS

Personal health record; P-sample; Q-sample; Q-statement; qualitative research; self-efficacy

XSL•FO
RenderX

Introduction

Recently, personal health records (PHRs) have been receiving increased attention from both the information and communications technology industry and academia as tools for consolidating, recording, and self-managing personal health information, as well as enabling self-efficacy, or the ability of patients to manage their own health. In the United States, PHRs such as Epic's MyChart, Dossia, and Microsoft's HealthVault that allow users to manage all of their health information in a single application are gaining many satisfied users, and many health care organizations have also built PHRs internally. According to the California Health Association's research, 1 out of 14 Americans use PHRs, and the number has doubled in the past 2 years. Also, a majority of the users are of high-income demographic status, and they tend to be younger than those who do not use PHRs. However, low-income and older patients are also increasingly adopting PHRs, as are those with chronic illnesses [1]. This may be because the system provides benefits for health management regardless of income or age. In one survey, two out of three responders expressed concerns about the security and privacy of their health information, but responders' concerns were reduced after experiencing the many benefits of PHRs. Additionally, it is predicted that the quality of life may even increase and the cost of health care will fall as PHRs become more widely used [2]. The primary participants in the PHR industry can be divided into service providers (such as the medical, health management, and information technology industries), consumers who use PHR services, and central and local governments that support the development and advance of the technology. To realize the goals of PHRs, the perspectives of all these participants need to be considered and fairly reflected [3].

The Healthcare Information Management and Systems Society has defined PHRs as follows [4]:

> *An electronic Personal Health Record ("PHR") is a universally accessible, layperson comprehensible, lifelong tool for managing relevant health information, promoting health maintenance and assisting with chronic disease management via an interactive, common data set of electronic health information and e-health tools. The ePHR is owned, managed, and shared by the individual or his or her legal proxy(s) and must be secure to protect the privacy and confidentiality of the health information it contains. It is not a legal record unless so defined and is subject to various legal limitations.*

The Medical Library Association/National Library of Medicine Joint Electronic Personal Health Record Task Force also examined the state of PHRs in an extensive review in 2010. After examining various existing definitions, they provide the following working definition [5]:

> *Electronic personal health record (PHR): a private, secure application through which an individual may access, manage, and share his or her health information. The PHR can include information that is entered by the consumer and/or data from other sources such as pharmacies, labs, and health care providers. The PHR may or may not include information from the electronic health record (EHR) that is maintained by the health care provider and is not synonymous with the EHR. PHR sponsors include vendors who may or may not charge a fee, health care organizations such as hospitals, health insurance companies, or employers.*

We recently published a systematic review on the history and trends of PHR research [6]. To assess the research efforts concerning the PHR to date, we searched the literature on research involving PHRs and have summarized the results, as well as describing how the topics assessed have evolved over time. For the search strategy, we queried PubMed, which returned 695 results. Through one-by-one analysis, we removed the results with the acronym PHR but with different definitions. In the end, we analyzed a total of 229 articles. The first appearance of PHR in an academic journal was in Germany in 1969—"Personal record linkage," in *Methods of Information in Medicine Supplement*. However, forms were, of course, not computerized at that time, so the early literature on the PHR refers to a simple collection of paper. In other words, the PHR in a historical context was a simple collection of notes containing information on one's health, and early studies of PHRs focused on such paper records. The shift to patient-centeredness was found afterward, and the "P" for personal in PHR was frequently used as an acronym for patient in the 1990s. Also, the phrase *personally controlled health records* strongly expresses the rights of control over one's personal records. A similar but not identical example of the use of P as an acronym for parent was published in 1993, in the phrase *parent-held record*. The PHR started to be accepted as a separate concept from the electronic medical record (EMR) with the use of phrases such as *personal medical record* (1995) and *computer-generated patient-held medical record* (1996). The distinction between digitized and paper records in the medical field began when computerized records became the standard, and the word *electronic* was added to PHR in order to distinguish it from past paper records. In the middle of the 21st century, as the discussion of electronic health records (EHRs) became increasingly common, the term *personal* was added to EHR. This is also the period when the phrases *personal health application, personal health information, personal health folder*, and *personal health record books* came into use. As privacy and security were stressed, PHR sometimes referred to *protected health records*.

Even though the beginning of PHR research goes as far back as the 1960s, it was followed by a period of little endeavor. In the 1960s, several studies of PHRs per year were published, and this trend remained consistent until the early 2000s, when the number rapidly increased. This trend is the result of the emergence of the patient-centered care paradigm and the acknowledgment of the PHR as an important means of patient safety and eHealth because the electronic PHR can be accessed digitally from anywhere and at anytime. Additionally, the advance of Internet and information technology has enabled various enhancements of PHR functionality and expansion of applications.

XSL•FO
RenderX

In 229 articles, we analyzed the research participants, methods, and target diseases of 172 articles with abstracts in this previously published study [6]. The effects of the PHR on disease and health management were the most frequent research topics, followed by the required features of the PHR. Additionally, some studies dealt with application analysis in public health, which was initially deemed a crucial function of the PHR. As the history of the PHR is relatively short, several articles addressed the predictions regarding the future direction and the implications of PHRs. Naturally, the PHR literature overlaps at times with that around EHRs and EMRs, and few articles have made an effort to clearly delineate their differences. Given the nature of PHRs, privacy and security issues are included frequently.

The most frequently used method for PHR research was the survey method. The second most frequently used method was to analyze and test the PHR, where the focus of the studies was to investigate the various perspectives of PHR users through interviews and focus groups. In terms of the PHR being a newly developed record of health management, there were studies on recommending the initial developmental directions. A large portion fell under the *other* category because there exists a large number of varied approaches in studying PHR, which reflects the absence of a unified approach.

As such, the understanding of PHRs may differ depending on the unique perspective of each academic institution, industry, and related field. Thus, there may still be a lack of consensus in understanding what a PHR is, both conceptually and as it can be practically instantiated. This suggests that the meaning of the PHR might benefit from study, with special focus on the expert opinions from those who are actively researching and developing PHRs. An accurate understanding of the perspective of PHR experts may be valuable in considering the developmental directions and potential utility of PHRs.

In studying a new concept with an incomplete definition, such as the PHR, it is important to conduct investigative research, but it is also necessary to try to describe the subject phenomenon from a unique perspective. William Stephenson suggested Q methodology as a means of dialectically compositing the tradition of opposition methodologies such as quantitative and qualitative research, objectivity and subjectivity research, explanation and understanding methods, naturalism and humanism, and positivism and antipositivism [7]. PHRs are at a stage of development and consensus establishment. This makes PHRs a suitable application for Q methodology, as its primary objectives are exploring new and unfamiliar phenomena and those that require further development. Through categorical analysis of PHR experts' opinions using this method, this research considers the future understanding of PHRs, as well as its current utility and further developmental directions.

Methods

Step 1: Selection of the Q-Sample

Q methodology is a research method used to study people's "subjectivity"—that is, their viewpoints [8]. To study participants' subjectivity, Q methodology uses self-referencing statements (Q-samples), which refer to phrases that project the responders' emotions or expectations instead of facts. A group of such phrases is referred to as a Q-population, and it is obtained through literature surveys and interviews regarding the research protocol. Hundreds of Q-populations are sampled by means of a literature survey and interviews, and Q-samples are selected by random and systematic sampling methods. In its first stage, this study sampled Q-populations regarding the PHR. Initially, broad literature surveys were used to collect diverse definitions and descriptions of PHRs, followed by consolidation of similar meanings and expressions. Excluding slight differences in expressions and word arrangements, the number of specific arrangements of words in definitions of PHRs available in the published literature is finite. This signifies mostly common opinions among scholars and experts regarding the major concepts. Among the available PHR Q-samples, we selected 50 Q-statements, which we divided into 5 categories: (1) characteristics, (2) functionality, (3) form, (4) requirements, and (5) business model. We selected these statements to ensure accuracy, maximize comprehensiveness, and include a variety of accurate positive, negative, and neutral statements. As a result, the Q-statements used herein consisted of 13 characteristic statements, 11 functionality statements, 11 form statements, 7 requirement statements, and 8 business model statements.

Step 2: Selection of the Person-Sample

Because Q methodology deals with differences in individual perspectives on relative importance, and not differences between individuals themselves, the number of person-samples (P-samples) included is not restricted. Rather, our research protocol is based on the small-sample doctrine [9]. Thompson [10] stated that opinions are best assessed through the following 5 groups: (1) those with special interest, (2) those who can judge and provide dispassionate interest, (3) those with authorities and expertise, (4) those with general interest but no special expertise, and, finally, (5) those who are uninformed and/or uninterested. The current stage of the research is not focused on all PHR users, but on the following 3 categories among the Thompson schemata who represented the P-samples: (1) special interest: PHR development executives and staff, research staff of PHR development projects; (2) authorities and experts: medical and health informatics professors, doctors, and nurses; and (3) class interests: medical and health informatics graduate students. We asked the Korean Society of Medical Informatics, which is the representative and the largest group for this professional discipline, to recommend experts for this domain, and we then invited those experts to participate voluntarily. Snowball sampling and personal contacts through professional networking were also used to reinforce the invitation of experts to form the valid P-samples. The institutional review board of the principal author's university reviewed and approved the research, and informed consent was collected from the participants.

Step 3: Q-Sorting

The Q-sorting stage of Q methodology requires researchers to arrange the P-sample statements into distributions according to individual degrees of agreement. In the present study, this

XSL•FO
RenderX

consisted of arranging the 50 Q-sample statements into priority groups with limits on how many statements can belong to each group, starting from the highest degree of agreement to the lowest. Statistically normal distribution was used as a forced distribution to specify the limits on the priority groups (Table 1). Q-samples in Q methodology represent a portion of subjective human opinion, and thus require a systematic forced distribution of the relative importance of statements instead of using an individual scoring system. The research was conducted from June 14 to 30, 2010, and the data were collected by individual interviews with P-samples, which included explanations of the objectives and methods of the research. In Q methodology, the participants are asked to provide further descriptions of the 2 statements with the largest opposing degree of agreement to aid further in Q-factor interpretation. This was included in the explanation during the interview, and investigative analysis was conducted accordingly.

Table 1. Distribution of Q-sorting

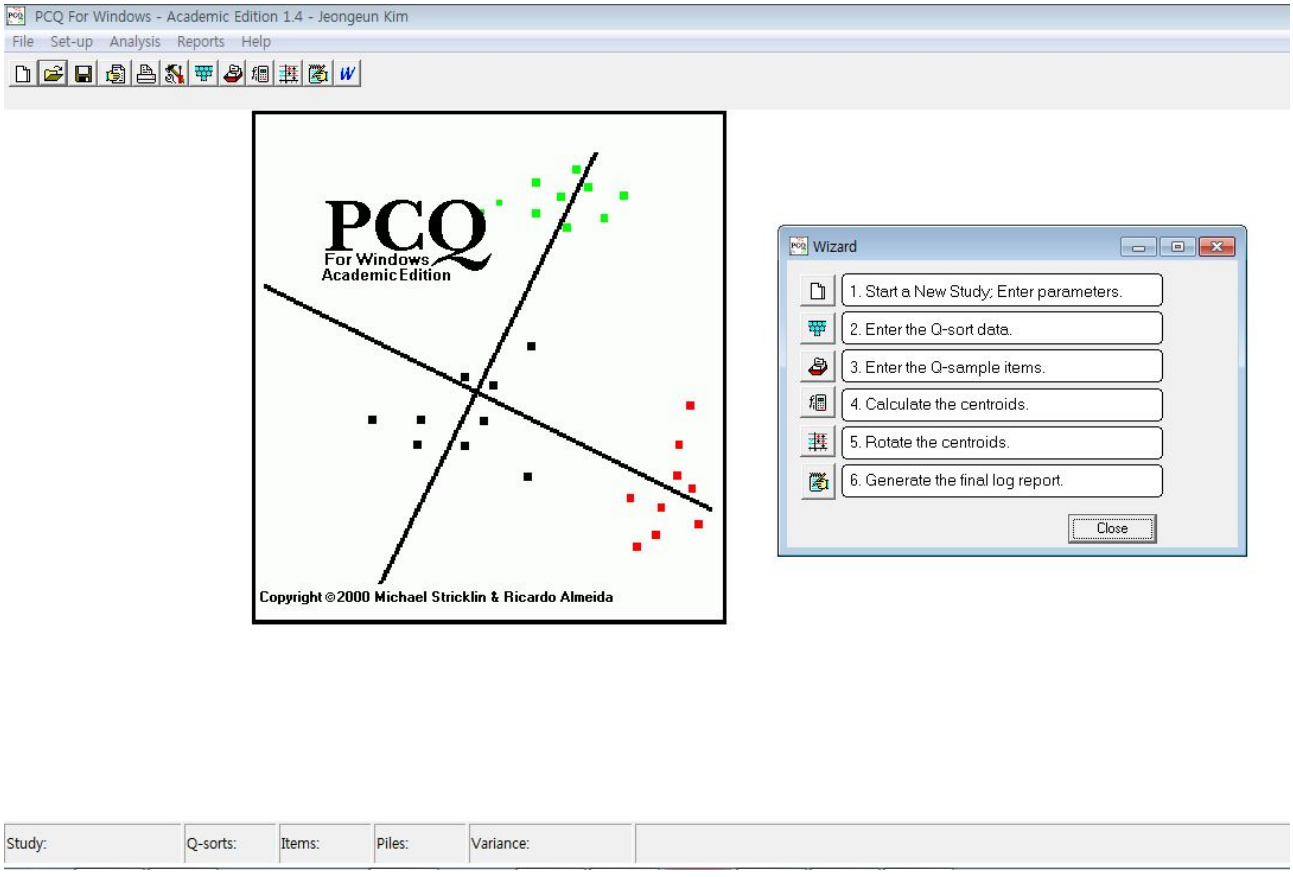
	Disagree			Neutral			Agree		
Score	–4	–3	–2	–1	0	1	2	3	4
Number of Q-samples[a]	2	4	6	8	10	8	6	4	2

[a] N (total number of Q-samples) = 50.

Statistical Analysis

Q methodology analysis is conducted through special software packages such as PCQ for Windows (PCQ Software, Portland, OR, USA), which we used in the present research (Figure 1). Figure 2 and Figure 3 show representative data layouts. In addition, we used SPSS version 19.0 (IBM Corporation, Somers, NY, USA) to reinforce the data interpretation.

Figure 1. PCQ for Windows.



XSL•FO
RenderX

Figure 2. Sample data entry layout for Q-sample items (English translation of items in Multimedia Appendix 1) in PCQ for Windows.

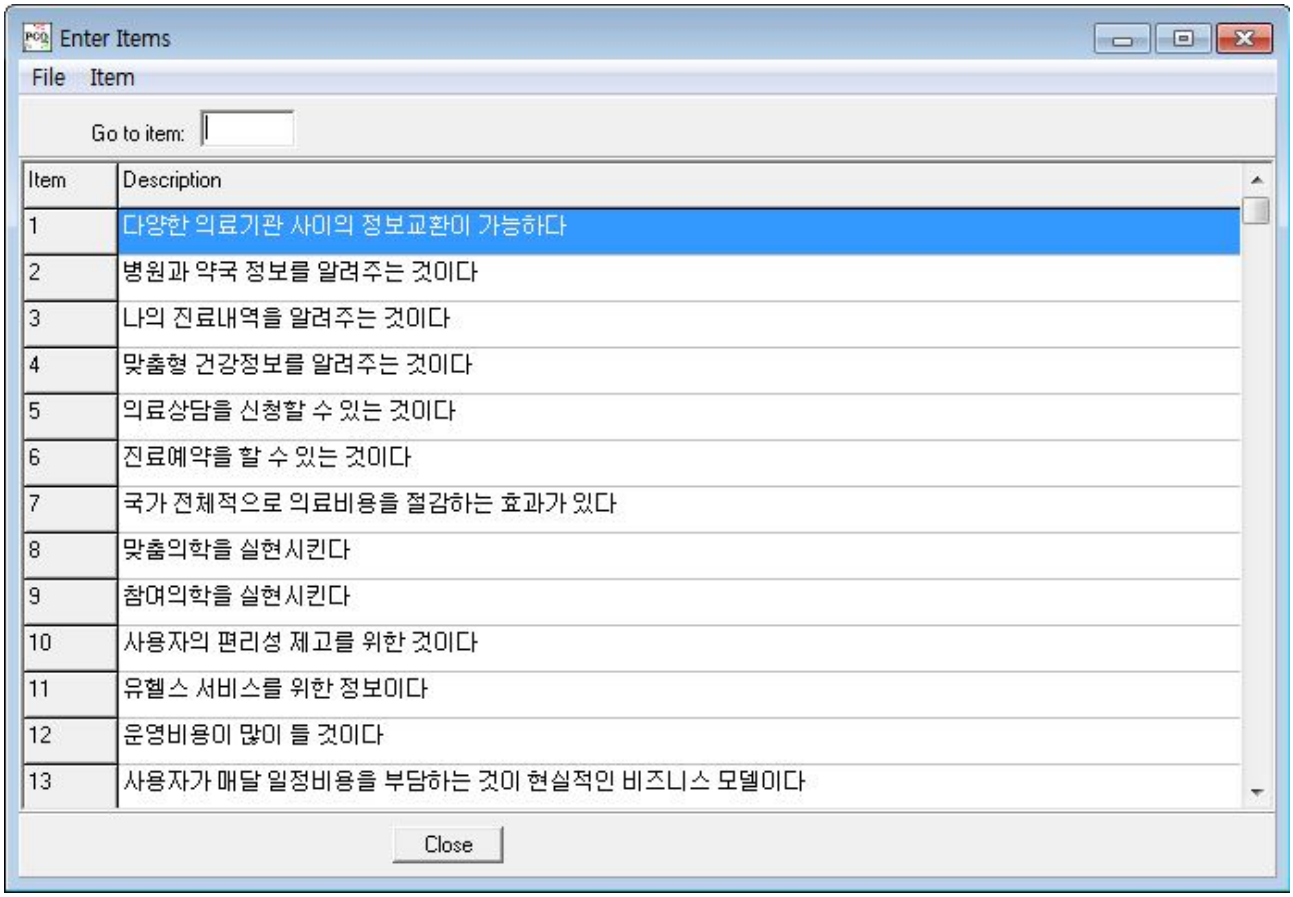
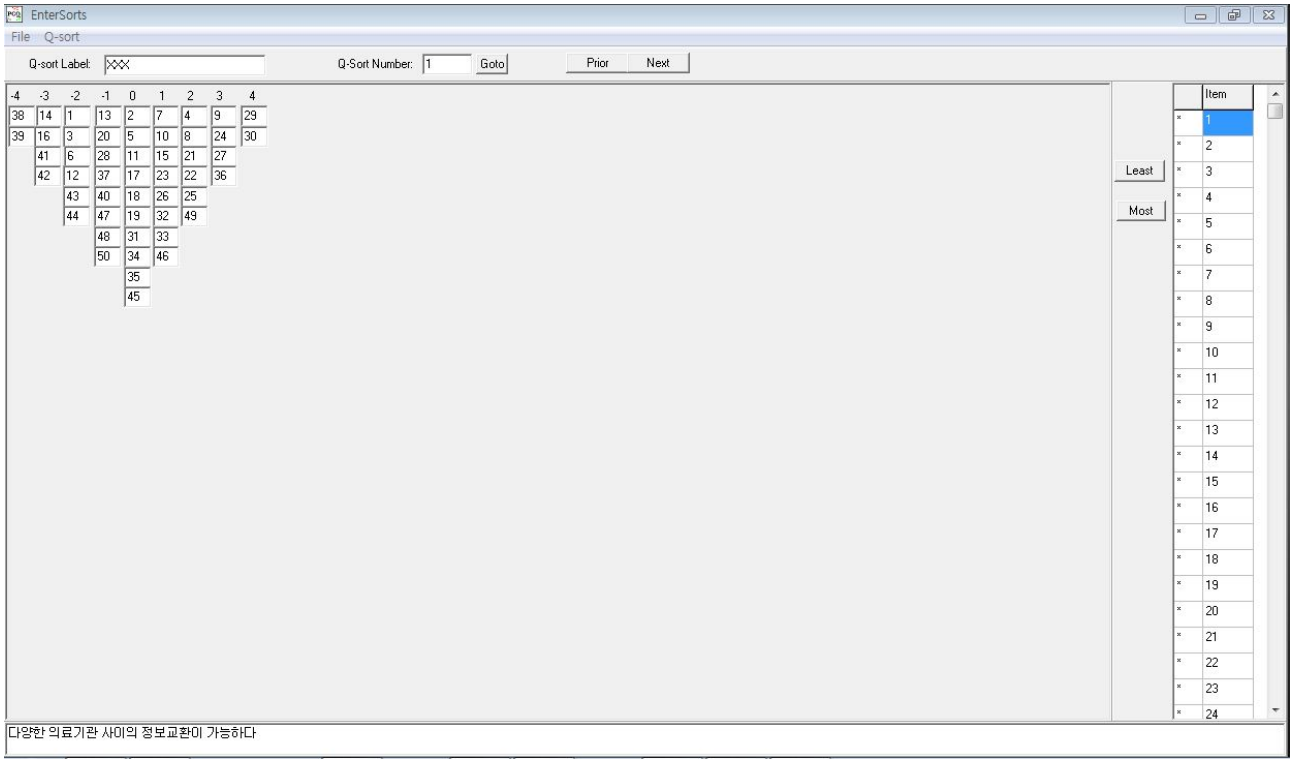


Figure 3. Sample data entry layout for the Q-sort data in PCQ for Windows.



XSL•FO
RenderX

Results

General Characteristics of Participants

We chose 45 participants for the study: 29 male and 16 female. Fields of expertise included 22 medical experts (9 doctors and 13 nurses) and 23 nonmedical experts (informatics engineering, computer engineering, genetics engineering, library and information science, etc). The average age of participants was 36.9 years (12 in their 20s, 16 in their 30s, 12 in their 40s, and 5 in their 50s).

Categorization of Experts' Opinions Regarding PHR According to Importance

Data were collected and analyzed using factor analysis in PCQ for Windows and SPSS 19.0. We used 45 Q-sorts as variables. The correlation coefficient was calculated and the correlation matrix was analyzed. In factor extraction, a larger-than-expected number of 18 factors had eigenvalues higher than 1 and, as it was impossible to apply these factors to factor rotation, extraction was based on the number of factors that was meaningful. The standards for selecting the number of factors were determined by inspecting the scree plot, which graphs the eigenvalue against the factor number for areas of sudden decrease in eigenvalues. Additionally, to determine the number logically, we compared and analyzed the results from setting the factor number to 5, 6, or 10, and we finally determined that the appropriate number was 5. The method used for factor extraction was principle component analysis, and the method of rotation was based on the results of processing through varimax rotation with Kaiser normalization repeated seven times. The resulting total variance of the 5 factors was 41.4%. On the basis of these analyses, we categorized expert opinions regarding PHR into 5 types (Table 2). The concepts that make up these 5 causes were organized and interpreted based on abduction—deriving a hypothesis from the observed facts—and descriptive processes.

Table 2. Personal health record statements with the strongest agreement and disagreement

Item number	Statement	Occurrences
Strongest agreement statements (+5)		
21	It is a lifetime health record of an individual	15
29	It is a realization of health 2.0 with the participation of clinical consumers	7
36	Its security is of utmost importance	7
33	It requires the protection of privacy regulation	6
34	It needs to be accessible anytime, anywhere	6
35	Its standardization is crucial	6
1	It is possible to exchange data among a variety of medical institutions	4
Strongest disagreement statements (–5)		
44	It is paper based	29
50	It is most effective to store it in portable USB memory sticks	8
24	It is not a legal document	6
39	It requires accreditation by the government	6

Type I: Public Interest Centered

The eigenvalue of type I was 6.5 and the variance percentage was 14.6%; 15 participants belonged to this group. This group considered the PHR to be a lifetime health record of an individual, that it requires the protection of privacy regulation, and that security is of the utmost importance. Paper-based and USB stick-based portability received low ratings, and whether the document is a legal document was questioned. Additionally, this group perceived the business model, in which the users pay a monthly fee, as impractical. In other words, this group regarded legislation and security as of the highest priority, and objected to models in which a significant burden was placed on individual users.

Type II: Health Information Standardization Centered

The eigenvalue of type II was 3.4 and the variance percentage was 7.5%; 7 participants belonged to this group. The group also agreed that the PHR should be a lifetime health record of an individual, and considered the standardization of the PHR crucial. This group also viewed the idea of paper- or USB stick-based portable PHRs unfavorably, in addition to having negative opinions regarding data exchange among a variety of medical institutions. In other words, they considered the one main functionality of the PHR to be information exchange, and because this is currently not realized, standardization needs to be an early focus of development.

Type III: Health Consumer Centered

The eigenvalue of type III was 2.9 and the variance percentage was 6.6%; 9 participants belonged to this group. This group considered the PHR to be a realization of health 2.0 with the participation of clinical consumers, and strongly sided with consumerism. This group did not consider the completeness of information, nor accreditation by the government, to be a high priority. Similar to the other groups, they did not favor the paper-based PHR, nor a business model based on advertisement revenue.

XSL•FO
RenderX

Type IV: Health Information Security Centered

The eigenvalue of type IV was 2.9 and the variance percentage of 6.5%; 7 participants belonged to this group. The group considered the PHR to be a lifetime health record of an individual that requires stringent security, but disagreed that it is not a legal document and should not be paper based. This concept of the PHR shows a similarity with type I, but with a lower rank.

Type V: Health Consumer Convenience Centered

The eigenvalue of type V was 2.7 with a variance percentage of 6.0%; 7 participants were included in this group. This group considered the PHR to be a realization of health 2.0 with the participation of clinical consumers, and they felt strongly that the PHR needs to be accessible anytime, anywhere. However, this group was similar to type IV, in that the group questioned whether accreditation by the government was crucial, and questioned the effectiveness of paper- or USB stick-based portability.

Consensus Regarding the Important Characteristics of PHR

The mixed research method of qualitative analysis along with quantitative methods, such as Q methodology, does not place great importance on the frequency, but rather on the weight, of meaning or relative relationships of the subject phenomena. In other words, a statement by itself has a meaning, but it also gains another meaning when relative comparison is made with other statements [11]. Therefore, the focus is not on how frequently the statement has been agreed upon, but rather on consolidating the statements that participants can commonly agree upon. Ultimately, current expert opinion of the PHR possesses the primary characteristics presented in Table 2.

Discussion

We found that there was broad agreement that the PHR should be a lifetime health record of an individual, and it primarily requires the participation of clinical consumers. Respondents felt that other approaches, such as placing the PHR on a USB, and requiring consumers to pay a monthly fee were less likely to be practical. As expected, the different groups had differing perspectives regarding which aspects of the PHR need attention most urgently, with the largest group focusing on public interests and smaller groups focusing on a health information-centered approach, a consumer-centered approach, security as the central concern, and consumer convenience as the primary issue.

The Q methodology used in this research highlights specific behaviors in a group or quantifies the minority groups, thereby reflecting the general behavior of a larger group by studying a fraction of the group. We hypothesized that many aspects of the PHR would be divided into diverse groups and aimed to extract the primary concepts. According to Brown, the Q-sort of 50 statements applied to 45 participants and in cause analysis with a characteristic coefficient larger than 1 are both sufficient for drawing a conclusion [9]. On analysis, the topic of discussion is the categorization of opinion types among the experts. This means that these types must be considered primary concepts with regard to user uptake and use in future research on and industrialization of PHR. Because of the nature of Q methodology, this research does not assert that the result is statistically proven, which would require additional research. The 5 types extracted in this research are not statistically confirmed but are identified as impressionistic conclusions. Other research employing Q methodology discusses similar points [12]. The types of expert opinions regarding PHR identified in this study require further supplementation and proof through research efforts with a separate methodology.

Conclusions

The PHR, which is appropriately receiving close attention from the medical and information technology industries, is likely to be widely adopted soon by large numbers of clinical consumers in developed countries. For the PHR to be efficiently used by the general public, an initial understanding of future developers' and users' opinions and preferences is required. Simultaneously, an accurate understanding and categorical analysis of opinions of those experts who lead the development and growth of PHR will be valuable to its adoption and expansion. In this research, we used Q methodology to categorize expert opinions on PHR. We identified 5 categories of perspectives centered on public interest, health information standardization, the health consumer, health information security, and health consumer convenience. Clearly, these are all important domains of the PHR that deserve attention. The medical industry should be developing detailed strategies for product development that address all these dimensions in order to win the support of people from all 5 perspectives. If PHRs are to achieve their considerable potential for improving health, they will need to contain sufficient content to be attractive to consumers, address their main concerns about areas such as security, and at the same time be based on business models that are successful in the long term. The domains that we identified are all going to continue to be important, but they will also evolve over time as PHRs evolve and grow more sophisticated.

The exact shape of future information technology applications is impossible to predict. Nonetheless, the PHR appears to be certain to have a key place at the table, since it will allow individuals to increase the quality of their lives by managing their own health information, a central point on which our participants agreed. The accurate understanding and categorical analysis of opinions of those experts who lead the development and growth of PHR presented in this study should inform the adoption and expansion of the PHR, thus ensuring its widespread uptake and clinical success.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST)(2011-0018261). The authors appreciate the enormous support for data collection and management effort of

XSL•FO
RenderX

Ms Kyungsoon Lee, who is studying for a Master's degree in Nursing Informatics at the College of Nursing, Seoul National University.

Conflicts of Interest

None declared

Multimedia Appendix 1

PHR Q-samples.

[PDF File (Adobe PDF File), 15KB - jmir_v13i4e105_app1.pdf]

References

1. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
2. KorMedi. KoreaMedicare. 2010 Jul 26. Increase of Satisfaction Towards the PHR (in Korean) URL: http://www.kormedi.com/news/article/1197188_2892.html [accessed 2010-07-26] [WebCite Cache ID 5rWfAXmDl]
3. National Information Society Agency. A Study on Market Research and Value Chain of u-Health Services. In: NIA II-RER-07075. Seoul, Korea: NISA; 2007.
4. HIMSS Board of Directors. Healthcare Information Management and Systems Society. 2007 Sep 28. HIMSS Personal Health Records Definition and Position Statement URL: http://www.himss.org/content/files/PHRDefinition071707.pdf [accessed 2010-12-17] [WebCite Cache ID 5v3H2FgeV]
5. Jones DA, Shipman JP, Plaut DA, Selden CR. Characteristics of personal health records: findings of the Medical Library Association/National Library of Medicine Joint Electronic Personal Health Record Task Force. J Med Libr Assoc 2010 Jul;98(3):243-249. [doi: 10.3163/1536-5050.98.3.013] [Medline: 20648259]
6. Kim J, Jung H, Bates DW. History and trends of "personal health record" research in PubMed. Healthc Inform Res 2011 Mar;17(1):3-17. [doi: 10.4258/hir.2011.17.1.3] [Medline: 21818452]
7. Stephenson W. Application of communication theory: the substructure of science. Psychol Rec 1972;22:17-36.
8. International Society for the Scientific Study of Subjectivity. 2011 Aug 04. Q Methodology URL: http://qmethod.org/about [accessed 2011-08-04] [WebCite Cache ID 60g674JTW]
9. Brown SR. Political Subjectivity: Applications of Q Methodology in Political Science. New Haven, CT: Yale University Press; 1980.
10. Thompson GC. The evaluation of public opinion. In: Berelson B, Janowitz M, editors. Reader in Public Opinion and Communication. 2nd edition. New York, NY: Free Press; 1966:7-12.
11. Schmolck P. University of the Federal Armed Forces Munich. 2010 May 02. The QMethod Page URL: http://www.lrz-muenchen.de/~schmolck/qmethod/ [WebCite Cache ID 5pRJnFkGy]
12. Xu CY, Kim KI. Study of cultural adaptation and attitude of Chinese students in Korea using Q method. Ment Health Res 1997;16:166-180.

Abbreviations

EHR: electronic health record
EMR: lectronic medical record
PHR: personal health record

Edited by G Eysenbach; submitted 25.02.11; peer-reviewed by D Jones, K Nazi; comments to author 27.07.11; revised version received 12.08.11; accepted 26.08.11; published 29.11.11.

Please cite as:
Kim J, Bates DW
Analysis of the Definition and Utility of Personal Health Records Using Q Methodology
J Med Internet Res 2011;13(4):e105
URL: http://www.jmir.org/2011/4/e105/
doi:10.2196/jmir.1781
PMID:22126860

XSL•FO
RenderX

©Jeongeun Kim, David W Bates. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Determinants of Engagement in Face-to-Face and Online Patient Support Groups

Cornelia F Van Uden-Kraan[1], PhD; Constance HC Drossaert[2], PhD; Erik Taal[2], PhD; Willem M Smit[3], MD, PhD; Hein J Bernelot Moens[4], MD, PhD; Mart AFJ Van de Laar[2,5], MD, PhD

[1]Department of Otolaryngology / Head and Neck Surgery, VU University Medical Centre, Amsterdam, Netherlands

[2]Institute for Behavioral Research, Department of Psychology, Health, Technology, University of Twente, Enschede, Netherlands

[3]Department of Internal Medicine, Medisch Spectrum Twente, Enschede, Netherlands

[4]Department of Rheumalotogy, Ziekenhuisgroep Twente, Hengelo, Netherlands

[5]Department of Rheumatology, Medisch Spectrum Twente, Enschede, Netherlands

Corresponding Author:
Constance HC Drossaert, PhD
Institute for Behavioral Research
Department of Psychology, Health, Technology
University of Twente
PO Box 217
Enschede, 7500 AE
Netherlands
Phone: 31 534896049
Fax: 31 534892388
Email: c.h.c.drossaert@utwente.nl

Abstract

Background: Although peer-to-peer contact might empower patients in various ways, studies show that only a few patients actually engage in support groups.

Objective: The objective of our study was to explore factors that facilitate or impede engagement in face-to-face and online peer support, using the Theory of Planned Behavior.

Methods: A questionnaire was completed by 679 patients being treated for arthritis, breast cancer, or fibromyalgia at two Dutch regional hospitals.

Results: Our results showed that only a minority of the patients engaged in organized forms of peer support. In total 10% (65/679) of the respondents had engaged in face-to-face meetings for patients in the past year. Only 4% (30/679) of the respondents had contact with peers via the Internet in the past year. Patients were more positive about face-to-face peer support than about online peer support ($P < .001$). In accordance with the Theory of Planned Behavior, having a more positive attitude ($P < .01$) and feeling more supported by people in the social environment ($P < .001$) increased the intention to participate in both kinds of peer support. In addition, perceived behavioral control ($P = .01$) influenced the intention to participate in online peer support. Nevertheless, the intention to engage in face-to-face and online peer support was only modestly predicted by the Theory of Planned Behavior variables ($R^2 = .33$ for face-to-face contact and $R^2 = .26$ for online contact).

Conclusion: Although Health 2.0 Internet technology has significantly increased opportunities for having contact with fellow patients, only a minority seem to be interested in organized forms of peer contact (either online or face-to-face). Patients seem somewhat more positive about face-to-face contact than about online contact.

(J Med Internet Res 2011;13(4):e106) doi:10.2196/jmir.1718

KEYWORDS

patients; support groups; online communities; breast cancer; fibromyalgia; rheumatoid arthritis; theory of planned behavior

XSL•FO
RenderX

Introduction

Over the past decades, several studies have shown that patients tend to profit from engagement in peer support groups. Such group participation can offer emotional support, confidence, and strength [1]; can foster hope [2]; and can lead to improved coping [3], less distress [4], and an improved quality of life for the participant [5]. Despite these empowering outcomes of engagement in peer support, studies have shown that many face-to-face peer support groups have only small numbers of participants [6,7].

People who engage in face-to-face peer support groups are more likely than nonparticipants to be female, younger, more highly educated, and of a higher economic status [8-10]. Contradictory findings appeared concerning social support: some studies found that those who participated in support groups experienced less social support in their social environment than nonparticipants [8-10], while others found no differences [6,11,12]. In general, participants seemed to be more anxious about their illness and had greater emotional problems than did nonparticipants [9,13].

With the availability of the Internet, so too the opportunity to share concerns and experiences with peers online has become available. The outcomes of participation in online support groups are in line with the outcomes of participation in face-to-face support groups [14,15]. Engagement in peer support was expected to increase with the emergence of online support groups, as these kinds of groups have specific advantages, such as the absence of geographical barriers, 24-hour availability, and anonymity [16,17]. Yet studies have shown that the use of online patient support groups is limited as well. For example, Atkinson et al [18] found that only 3.8% of their sample of Internet users had ever used an online patient support group, and van de Poll-Franse and van Eenbergen [19] found that only 6% of their sample of cancer patients had actually participated in an online peer support group.

Little is known about determinants of (non)participation in online support groups. We are aware of only two studies that provided some insight. Dutta and Feng [20] showed that participants are younger than nonparticipants. Atkinson et al [18] found that having a poorer health status and a lower income significantly increased use of online support groups for people with similar health or medical issues, while having access to the Internet both at home and at work significantly decreased the use.

In the present study we focused on determinants of engagement in online as well as face-to-face peer support. It is important to gain more insight into the factors that impede or facilitate engagement in peer support, because the numbers of patients that benefit from it might be increased when misconceptions of and barriers to peer support are removed.

As the theoretical basis for the present study, we chose the Theory of Planned Behavior [21]. According to the Theory of Planned Behavior, intention to engage in peer support is determined by three considerations: (1) attitude (ie, thoughts and feelings regarding engagement in peer support), (2) subjective norm (ie, patients' perceptions about whether significant others would like them to engage in peer support), and (3) perceived behavioral control (ie, the extent to which patients think that they are able to engage in peer support).

The purpose of this quantitative study was (1) to explore differences in intentions, attitudes, social norms, and perceived behavioral control regarding face-to-face and online support groups, and (2) to examine which factors of Theory of Planned Behavior variables, demographic variables, health-related quality of life, and social support predict patients' intention to engage in both types of peer support. In this respect we were interested not only in which type of peer support could be better explained, but also in whether face-to-face and online peer support would have the same or differing predictors.

Methods

Sample and Procedure

Our study was focused on patients with breast cancer, fibromyalgia, or rheumatoid arthritis. We randomly selected 400 patients from each patient group from the electronic database of two regional hospitals. Inclusion criteria were being younger than 75 years and having sufficient knowledge of the Dutch language to be able to fill out the questionnaire. Attending physicians (n = 22) were asked to exclude those patients who did not meet the inclusion criteria and those who for other reasons were deemed unsuitable for participation in our study. Reasons mentioned for exclusion by the physicians were as follows: deceased, aggravation of the illness, comorbidity, mental health problems, wrong diagnosis, or family circumstances. Of the 22 physicians, 2 did not respond, which meant that 30 breast cancer patients were not approached. After exclusion of in total 187 patients we were left with a group of 1013 patients. The attending physicians invited the patients by mail and enclosed the questionnaire. If necessary, this was followed by one reminder. Of the 1013 patients approached, 28 were ineligible because they were deceased or had no valid address. The overall total response rate was 68.9% (n = 679). Of these respondents, 272/350 patients had a diagnosis of rheumatoid arthritis (response rate: 77.7%), 212/333 of fibromyalgia (response rate: 63.7%), and 195/302 of breast cancer (response rate: 64.6%).

All patients were asked for their consent to check the actual date of diagnosis in their medical records. According to the Dutch law for medical research with humans (Wet Medisch-wetenschappelijk onderzoek met Mensen), approval by an ethics committee was not necessary for this survey study.

Instrument

Demographic and Health Characteristics

The respondents were asked to provide information about the demographic characteristics sex, age, marital status, education, and employment. Health-related quality of life was assessed with the SF-12v2. Standardized scores were calculated for the physical and mental well-being varying from 0 (poor) to 100 (excellent), with a mean of 50 and a standard deviation of 10 in the general population of the United States [22].

XSL•FO
RenderX

Social Support Factors

Social support factors were measured by the Social Support List-Interaction [23], consisting of 12 items. An example of a social support item is "Does it ever happen that someone shows interest in you?" Respondents could answer on a 4-point scale that ranged from "seldom to never" (score of 1) to "often" (4). The internal consistency (Cronbach alpha) for this construct was alpha = .93. A mean total score was calculated.

Use of Peer Support

Respondents were asked to indicate whether they had contact with peers at patient meetings, via the Internet, or at patient organization venues, or whether they had contact with an acquainted peer during the past year and, if so, how frequent this contact was. Respondents could answer on a 4-point scale that ranged from "never" (1) to "regularly" (4).

Theory of Planned Behavior Variables

Theory of Planned Behavior variables were measured regarding both face-to-face and online peer contact. Items were derived from the literature (eg, [9,13,14, 24]). For each construct the internal consistency (Cronbach alpha) was determined and a mean total score was calculated.

We asked about patients' *intention* to have contact with peers face-to-face and via the Internet during the coming year on a 5-point scale that ranged from "certainly no" (1) to "certainly yes" (5). *Attitude* toward face-to-face and online peer support was measured directly with two single items: "Face-to-face contact with peers is valuable" and "Contact with peers via the Internet is valuable." Attitude was also measured indirectly by assessing advantages and disadvantages. In total, 28 items were formulated (see table 4). Respondents could answer on a 5-point scale that ranged from "totally disagree" (1) to "totally agree" (5). *Advantages of face-to-face peer support* was measured with 9 items (alpha = .93). *Disadvantages of face-to-face peer support* was measured with 5 items (alpha = .74). *Advantages of online peer support* was measured with 9 items (alpha = .92). *Disadvantages of online peer support* was measured with 5 items (alpha = .65).

Subjective norm was measured with two items: "People who are important to me think that I certainly should be in contact with peers face-to-face" and "People who are important to me think that I certainly should be in contact with peers via the Internet." Respondents could answer on a 5-point scale that ranged from "should not" (1) to "should" (5).

Perceived behavioral control was measured directly with two items: "I consider myself capable of having contact with peers face-to-face" and "I consider myself capable of having contact with peers via the Internet." Response options ranged from "totally disagree" (1) to "totally agree" (5). Perceived behavioral control was also measured indirectly by assessing *barriers*. In total, 13 items were formulated (see table 5). Respondents could answer on a 5-point scale that ranged from "very easy" (1) to "very difficult" (5). The barriers we asked about for face-to-face and online peer support partially differed, as a result of different characteristics. *Barriers to face-to-face peer support* was measured with 5 items (alpha = .83). *Barriers to online peer support* was measured with 8 items (alpha = .90).

Data Analysis

Differences in Theory of Planned Behavior variables concerning face-to-face and online peer support were tested by means of paired-sample *t* tests. We used a hierarchical multiple regression analysis to determine to what extent intention to engage in peer support could be predicted. The determinants of the Theory of Planned Behavior were entered in the first block of the regression analysis. In the second block social support factors, health-related characteristics, and the demographic characteristics that correlated significantly with intention were entered. Statistical significance was assumed when $P < .05$.

Results

Participants' Demographic and Health Characteristics

Most of the respondents were female (84.3%) (Table 1). The mean age of the respondents was 54 years. The majority of the respondents were married or living with a partner, had a low level of education, and were unemployed. Patients had a diagnosis of rheumatoid arthritis (40.1%), fibromyalgia (31.2%), or breast cancer (28.7%). The mean duration of the participants' illness was 7 years, with a range from 0 to 59 years.

The respondents had an average score of 38.6 on the physical component and an average score of 43.9 on the mental component of the SF-12v2. This indicates that the respondents' physical and mental well-being was worse than the average of the general population.

XSL•FO
RenderX

Table 1. Demographic and health characteristics of the participants and social support factors (602 ≤ n ≤ 679)

Sex, n (%)	
Female	571 (84.3%)
Male	106 (15.7%)
Age (years)	
Mean (SD)	54 (12.9)
Minimum	18
Maximum	75
Marital status, n (%)	
Single	128 (19.5%)
Married/cohabiting	530 (80.5%)
Education, n (%)	
Low	404 (59.9%)
Middle	176 (26.1%)
High	94 (13.9%)
Employment, n (%)	
Employed	212 (32.2%)
Unemployed	447 (67.8%)
Diagnosis, n (%)	
Breast cancer	195 (28.7%)
Fibromyalgia	212 (31.2%)
Rheumatoid arthritis	272 (40.1%)
Disease duration (years)	
Mean (SD)	7.1 (7.8)
Minimum	0
Maximum	59
Well-being (SF-12v2), mean (SD)	
Physical well-being	38.6 (11.3)
Mental well-being	43.9 (6.7)
Social support (score 1–4)	2.6 (0.66)

Use of Face-to Face and Online Peer Support

The majority of the respondents (n = 396, 58.3%) had contact with peers during the past year (data not in table). The most regular type of peer support was contact with an acquainted peer (353/679, 52.0%) (Table 2). In total, 9.6% (65/679) of the respondents had engaged in face-to-face meetings for patients in the past year. Only 4.4% (30/679) of the respondents had contact with peers via the Internet in the past year. Of the respondents, 5.3% (36/679) indicated they had contact in the past year with peers at patient organization venues.

Table 2. Use of peer support (n = 679) during the preceding year

Type of peer support	Never		Once		Several times		Regularly	
	n	%	n	%	n	%	n	%
How often did you have contact with peers at patient meetings during the past year?	614	90.4	26	3.8	26	3.8	13	1.9
How often did you have contact with peers via the Internet during the past year?	649	95.6	7	1.0	17	2.5	6	0.9
How often did you have contact with (an) acquaintance(s) with the same disease during the past year?	326	48.0	35	5.2	216	31.8	102	15.0
How often did you have contact with peers at patient organization venues during the past year?	643	94.7	15	2.2	12	1.8	9	1.3

XSL•FO
RenderX

Determinants of Theory of Planned Behavior Concerning Face-to-Face and Online Peer Support

The respondents' intention to engage in face-to-face and online peer support in the coming year was slightly negative (Table 3). Only a minority of the respondents intended to look for peers via the Internet (35/654, 5.4% [certainly] yes; 135/654, 20.6% maybe; 484/654, 74.0% [certainly] not) or for face-to-face peer contact (104/663, 15.7% [certainly] yes; 164/663, 24.7% maybe; 395/663, 59.6% [certainly] not) in the coming year (data not in table).

Table 3. Mean scores (range 1–5) for determinants of the theory of planned behavior toward face-to-face and online peer support

Determinant	Face-to-face peer support (601 ≤ n ≤ 663)		Online peer support (530 ≤ n ≤ 654)	
	Mean	SD	Mean	SD
Intention[a]	2.4	1.2	2.0	0.95
Attitude[a]	3.7	1.1	3.2	1.1
Advantages[a]	3.5	0.85	3.2	0.86
Disadvantages[a]	2.9	0.88	3.2	0.85
Subjective norm[a]	3.0	0.69	2.8	0.67
Perceived behavioral control[a]	4.1	1.2	3.9	1.4
Barriers[b]	2.8	0.80	2.8	0.91

[a] $P < .001$ for paired-sample t tests comparing face-to-face versus online peer support.

[b] No differences in amount of barriers between face-to-face and online peer support could be determined, because the questionnaire asked about different barriers.

Although the respondents had a slightly positive attitude toward both kinds of peer support, they were significantly ($P < .001$) more positive toward face-to-face support. Respondents experienced significantly greater advantages and fewer disadvantages using face-to-face support than using online peer support.

The scores on the separate items (Table 4) revealed that the most important advantages of both types of peer support were "sharing experiences" and "finding recognition."

Table 4. Mean item scores (range 1–5) on attitude toward peer support

	Face-to-face peer support (601 ≤ n ≤ 616)		Online peer support (526 ≤ n ≤ 546)	
	Mean	SD	Mean	SD
Advantages: (Through) contact with peers...				
Offers a good opportunity to share your experiences	3.8	1.1	3.6	1.1
provides recognition and understanding	3.8	1.1	3.6	1.0
provides support	3.7	1.1	3.4	1.0
is informative	3.5	1.1	3.4	1.0
is comforting	3.4	1.0	3.1	1.0
you feel empowered as a patient	3.4	1.1	3.1	1.1
provides reliable information	3.3	1.1	3.1	1.0
you can cope better with your illness	3.2	1.1	3.0	1.1
you can accept your illness more easily	3.2	1.1	2.9	1.1
Disadvantages : (Through) contact with peers...				
you are occupied too much with your illness	3.2	1.4	3.6	1.3
is too informal	2.9	1.1	3.2	1.0
makes people more concerned about the consequences of their disease	2.9	1.2	3.1	1.2
is too shallow	2.7	1.1	3.2	1.0
takes too much time	2.7	1.3	2.9	1.3

XSL•FO
RenderX

The most important disadvantage of both types of peer support was the continual confrontation with their illness. In general, respondents felt significantly more encouraged by people in their social environment to be in contact with peers face-to-face than via the Internet. The respondents considered themselves significantly more capable of having contact with peers face-to-face than via the Internet.

Differences in perceived barriers between face-to-face and online peer support could not be determined, because different barriers were asked about. The scores on the various barriers (Table 5) revealed that for both kinds of peer support, the most important one was to find a suitable peer support group. For online peer support other important barriers were to discuss the illness on the Internet, to actually write about the illness on the Internet, and the difficulty (due to the illness) of having to type or sit behind the computer for a long period of time.

Table 5. Mean item scores (range 1–5) for barriers to peer support

How difficult or how easy is it for you...	Mean	SD
Face-to-face peer support (582 ≤ n ≤ 595)		
to find a suitable face-to-face peer support group?	3.0	1.0
to find the time and the opportunity to contact peers face-to-face?	2.9	0.97
considering your illness to visit face-to-face peer support groups?	2.9	1.0
to afford the cost involved with face-to-face peer contact?	2.8	1.0
to talk about your illness with peers face-to-face?	2.7	1.0
Online peer support (506 ≤ n ≤ 518)		
to find a suitable online peer support group?	3.1	1.1
to talk about your illness on the Internet?	3.1	1.1
to verbally express your illness on the Internet?	3.1	1.1
considering your illness to type or sit behind the computer for a long period of time?	3.1	1.2
to find the time and the opportunity to contact peers via the Internet?	3.0	1.0
to afford the costs involved with peer-to-peer contact via the Internet?	2.6	1.1
to work with the Internet?	2.3	1.3
to obtain access to the Internet?	2.2	1.2

Prediction of Intention

Theory of Planned Behavior variables explained 33.3% of the intention to engage in face-to-face contact. Of the distal factors, physical and mental well-being, sex, and past behavior significantly improved the total amount of explained variance of intention to engage in face-to-face support. After inclusion of these distal factors, the influence of TBP variables remained significant (Table 6).

Theory of Planned Behavior variables explained 26.3% of the intention to engage in online contact. Of the distal factors, mental health, age, and past behavior significantly improved the total amount of explained variance of intention to engage in online peer support. The influence of Theory of Planned Behavior variables on intention remained significant after inclusion of the distal factors.

The total amounts of explained variance were moderate for face-to-face contact (40.0%) and online contact (36.2%).

When repeating the analysis among only those patients who had not had (online) contact with fellow patients in the past, we found similar results: Theory of Planned Behavior variables explained 27.5% of intentions to engage in face-to-face support and 24.6% of online peer support (data not shown).

XSL•FO
RenderX

Table 6. Extent to which intention to seek online peer support and face-to-face peer support can be predicted by determinants of the theory of planned behavior

Determinant	Intention to seek face-to-face peer support (n = 554)			Intention to seek online peer support (n = 489)		
	r	Beta	SE	*r*	Beta	SE
Step 1						
Attitude		0.22***	0.05		0.19***	0.05
Advantages		0.17**	0.05		0.20***	0.05
Disadvantages		–0.19***	0.04		–0.11**	0.04
Subjective norm		0.22***	0.04		0.19***	0.04
Perceived behavioral control		0.03	0.04		0.16**	0.05
Barriers		–0.02	0.04		0.12*	0.05
	R^2 = .33, $F_{6,548}$ = 45.1***			R^2 = .26, $F_{6,483}$ = 28.7***		
Step 2						
Attitude	.46**	.19***	0.05	.38**	.15**	0.05
Advantages	.46**	.17**	0.05	.40**	.18***	0.05
Disadvantages	–.33**	–.19***	0.04	–.19**	–.10*	0.04
Subjective norm	.37**	.19***	0.04	.30**	.17***	0.04
Perceived behavioral control	.19**	.00	0.03	.24**	.11*	0.04
Barriers	–.18**	–.09*	0.04	–.07	.11*	0.05
Social support factors	.02	–.06	0.04	–.02	–.01	0.04
Physical health	–.14**	–.13**	0.04	–.23**	–.08	0.04
Mental health	–.10**	–.09*	0.04	–.23**	–.12**	0.04
Time since diagnosis	–.06	.07	0.04	–.14*	–.02	0.04
Arthritis versus breast cancer	–.02	–.08	0.05	–.21**	–.05	0.05
Arthritis versus fibromyalgia	.13**	.01	0.05	.30**	.03	0.06
Sex (male vs female)	.11**	.08*	0.04	.03	–.04	0.04
Age (years)	–.15**	.00	0.05	–.34**	–.14**	0.05
Marital status (married/cohabiting vs single)	.07	.02	0.04	.09*	.00	0.04
Support group past behavior	.30**	.15***	0.04	.27**	.13**	0.04
	R^2 = .40, $F_{16,538}$ = 22.1***			R^2 = .36, $F_{16,473}$ = 16.8***		
R^2 change	.07***			.10***		

* $P < .05$, ** $P < .01$, *** $P < .001$.

Discussion

To the best of our knowledge, this study is the first to examine which psychological determinants predict patients' intention to engage in face-to-face and online peer support. Earlier studies focused only on determinants of patients' intention to engage in face-to-face peer support, and frequently lacked a theoretical framework. Our study confirmed that only a relatively small percentage of the patients engaged in organized forms of peer support. The respondents were more positive about and more inclined to use face-to-face peer support than online peer support.

Our results are in contrast to our expectations, as we had expected that the Internet and Health 2.0 technology would significantly facilitate peer contact between patients. In the literature, many advantages of online support groups are mentioned, such as easy accessibility, no physical or geographic barriers, and 24-hour availability. An explanation might be found in the fact that we questioned a somewhat older patient population. Older people mostly treat the Internet with greater skepticism than do younger people. "Trust" is of specific importance to patient support groups, because the topic of "an illness" in itself requires a high level thereof, thus this might have influenced patients' perceptions of online peer support [25]. Another explanation may be that we only included patients with common diseases. For people who have a relatively rare

XSL•FO
RenderX

disease, online peer support can provide a particularly valuable alternative, because for them it is more difficult to find peers with the same or similar conditions with whom they can share their experiences near their local communities [26].

Our study revealed that in accordance with the Theory of Planned Behavior, having a more positive attitude, feeling more supported by people in the social environment, and feeling more able to participate in peer support increased the intention to participate in organized forms of peer support. However, it should be notified that perceived behavioral control is not significant for face-to-face support. This is in line with Grande et al [9], who found that a more positive attitude and a higher subjective norm increased engagement in (face-to-face) peer support. It is also in line with Voerman et al [10], who found that a more positive attitude and a higher perceived control increased intention to engage in peer support.

Intention to engage in peer support was only modestly predicted by the Theory of Planned Behavior variables (face-to-face: 33.3%; online 26.3%). A meta-analysis has shown that Theory of Planned Behavior variables, on average, account for 35%–50% of the variance in intention [27]. It is difficult to compare the amount of explained variance with results of others studying participation behavior in face-to-face peer support, because in these studies logistic regression analysis was used [9,10]. An explanation for the relatively low amount of explained variance might be that, although respondents thought that peer support was valuable, they did not consider it valuable for themselves personally. According to the Theory of Planned Behavior, people need to perceive benefits of engagement in peer support to be of personal importance, instead of only for others, if they intend to execute the examined behavior [9,21]. In addition, future research might benefit from a combination of theoretical models to explain engagement in peer contact. In particular, the social comparison theory [28] has been used previously to study effects of peer contact, and could also be valuable in examining patients' reasons for (not) participating in this type of contact. According to the social comparison theory, people have a drive to compare themselves with others who face similar challenges [28]. For patients this can lead to feeling "less alone" in coping with the disease [29]. In addition, upward social comparison (looking at people who are doing better) can be a source of inspiration and advice [30], while downward social comparison (looking at people who are doing worse) can lead to positive affect by providing examples of how bad things could be [31]. Although some of the assessed advantages and disadvantages in our measures did derive from social comparison theory, future studies could gain by more explicitly combining the two models.

Patients who indicated having poorer mental well-being had a greater intention to participate in face-to-face and online peer support, and those who had worse physical well-being were more inclined to participate in face-to-face peer support. These results are not surprising, considering that health-related support groups have a health-promotional function. Therefore, these groups are less appealing to patients who perceive themselves already having good mental and physical well-being despite their illness [18,32].

Of the demographic factors, only age significantly improved the total amount of explained variance of intention to engage in online peer support. Younger patients were more inclined to engage in online support groups. These results were in line with our expectations, as it is still mainly younger people who use the Internet [19].

Pointers for an Intervention

This study yielded some pointers for an intervention, so that patients can make well-informed decisions about whether they want to engage in peer support and so that they can find a peer support group the moment they want to enroll. First, attention should be paid to awareness of peer support. Our study revealed that a considerable proportion of patients expected difficulties with finding relevant peer groups, especially on the Internet. Since studies showed that not all people have the necessary Internet skills to be capable of finding the information and applications they are looking for [33], it can be expected that not all patients manage to find online peer support groups by themselves. Second, our study revealed that many potential participants perceived various disadvantages to peer support. A major concern is the confrontation with negative sides of the disease. In line with Winefield et al [13], we believe that an intervention should inform potential participants of the specific aim of peer support groups and how they operate. Patients could, for example, be encouraged to read along with an online peer support group (ie, so-called lurking). By lurking, patients get a feeling for how such a group operates and what kind of people participate [34]. In addition, it could be emphasized that an increasing number of online peer support groups also offer the opportunity for "buddy matching." An optimal peer match can have a positive influence on interpersonal trust, and this is an important basis for the exchange of experiences and empathy [35].

Limitations

The findings of this study are limited by its cross-sectional nature. Therefore, we could attribute no causal relationships.

A second limitation of this study is the high number of missing variables in the section of the questionnaire on TBP items concerning online peer support. A considerable number of people without computer skills did not respond to these items. In addition, findings for *disadvantages of online peer support* might be less reliable because of the relatively low alpha (alpha = .65) of this construct.

Conclusions

Although opportunities for having contact with fellow patients have been significantly increased by Health 2.0 Internet technology, only a minority of patients seem to be interested in organized forms of peer contact (either online or face-to-face). Patients seem somewhat more positive about face-to-face contact than about online contact.

Our study revealed that in accordance with the Theory of Planned Behavior, having a more positive attitude and feeling more supported by people in the social environment increased the intention to participate in both kinds of peer support. In addition, perceived behavioral control influenced the intention

XSL•FO
RenderX

to participate in online peer support. Nevertheless, we must conclude that the Theory of Planned Behavior variables only modestly predicted the intention to engage in face-to-face and online peer support.

Acknowledgments

This study was supported by grants from the Netherlands Organization for Health Research and Development (ZonMw), the Dutch Arthritis Association, and Comprehensive Cancer Center Netherlands (IKNL). We would like to thank the rheumatologists, medical oncologists, and radiotherapists of Medisch Spectrum Twente and the rheumatologists of the Ziekenhuisgroep Twente for their help in recruiting patients.

Conflicts of Interest

None declared

References

1. Cope DG. Functions of a breast cancer support group as perceived by the participants: an ethnographic study. Cancer Nurs 1995 Dec;18(6):472-478. [Medline: 8564944]
2. Vugia HD. Support groups in oncology: building hope through the human bond. J Psychosoc Oncol 1991;9(3):89-107. [doi: 10.1300/J077v09n03_08]
3. Spiegel D, Bloom JR, Yalom I. Group support for patients with metastatic cancer. A randomized outcome study. Arch Gen Psychiatry 1981 May;38(5):527-533. [Medline: 7235853]
4. Silverman-Dresner T. Self-help groups for women who have had breast cancer. Imagin Cogn Pers 1990;9(3):237-243.
5. van den Borne HW, Pruyn JFA, van Dam-de Mey K. Self-help in cancer patients: a review of studies on the effects of contacts between fellow-patients. Patient Educ Couns 1986;8(4):367-385. [doi: 10.1016/0738-3991(86)90103-5]
6. Hammann HA, Croyle RT, Smith KR, Quan BNM, McCane KL, Botkin JR. Interest in a support group among individuals tested for a BRCA1 gene mutation. J Psychosoc Oncol 2000;18(4):15-37. [doi: 10.1300/J077v18n04_02]
7. Taylor SE, Falke RL, Shoptaw SJ, Lichtman RR. Social support, support groups, and the cancer patient. J Consult Clin Psychol 1986 Oct;54(5):608-615. [Medline: 3771878]
8. Bauman LJ, Gervey R, Siegel K. Factors associated with cancer patients' participation in support groups. J Psychosoc Oncol 1992;10(3):1-20.
9. Grande GE, Myers LB, Sutton SR. How do patients who participate in cancer support groups differ from those who do not? Psychooncology 2006 Apr;15(4):321-334. [doi: 10.1002/pon.956] [Medline: 16106472]
10. Voerman B, Visser A, Fischer M, Garssen B, van Andel G, Bensing J. Determinants of participation in social support groups for prostate cancer patients. Psychooncology 2007 Dec;16(12):1092-1099. [doi: 10.1002/pon.1160] [Medline: 17278186]
11. Bui LL, Last L, Bradley H, Law CH, Maier BA, Smith AJ. Interest and participation in support group programs among patients with colorectal cancer. Cancer Nurs 2002 Apr;25(2):150-157. [Medline: 11984103]
12. Cameron LD, Booth RJ, Schlatter M, Ziginskas D, Harman JE, Benson SR. Cognitive and affective determinants of decisions to attend a group psychosocial support program for women with breast cancer. Psychosom Med 2005;67(4):584-589 [FREE Full text] [doi: 10.1097/01.psy.0000170834.54970.f5] [Medline: 16046371]
13. Winefield HR, Coventry BJ, Lewis MA, Harvey EJ. Attitudes of patients with breast cancer toward support groups. J Psychosoc Oncol 2003;21(2):39-54. [doi: 10.1300/J077v21n02_03]
14. van Uden-Kraan CF, Drossaert CH, Taal E, Shaw BR, Seydel ER, van de Laar MA. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008 Mar;18(3):405-417. [doi: 10.1177/1049732307313429] [Medline: 18235163]
15. Bartlett YK, Coulson NS. An investigation into the empowerment effects of using online support groups and how this affects health professional/patient communication. Patient Educ Couns 2011 Apr;83(1):113-119. [doi: 10.1016/j.pec.2010.05.029] [Medline: 20599338]
16. Finn J. An exploration of helping processes in an online self-help group focusing on issues of disability. Health Soc Work 1999 Aug;24(3):220-231. [Medline: 10505283]
17. Finfgeld DL. Therapeutic groups online: the good, the bad, and the unknown. Issues Ment Health Nurs 2000;21(3):241-255. [Medline: 11075065]
18. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
19. van de Poll-Franse LV, van Eenbergen MC. Internet use by cancer survivors: current use and future wishes. Support Care Cancer 2008 Oct;16(10):1189-1195. [doi: 10.1007/s00520-008-0419-z] [Medline: 18293014]
20. Dutta MJ, Feng H. Health orientation and disease state as predictors of online health support group use. Health Commun 2007;22(2):181-189. [doi: 10.1080/10410230701310323] [Medline: 17668997]

XSL•FO
RenderX

21. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
22. Ware JE, Kosinski M, Turner-Bowker DM, Gandek B. How to Score Version 2 of the SF-12® Health Survey. Lincoln, RI: QualityMetric Incorporated; 2002.
23. van Sonderen FLP. Het Meten van Sociale Steun met de Sociale Steun Lijst-Interacties (SSL-I) en Sociale Steun Lijst Discrepanties (SSL-D): Een Handleiding. Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken, Rijksuniversiteit Groningen; 1993.
24. van den Borne HW, Pruyn JFA. Informatiebehoefte en lotgenotencontact bij kankerpatiënten. Gezondheid en Samenleving 1984;5:180-187.
25. Leimeister JM, Ebner W, Krcmar H. Design, implementation and evaluation of trust-supporting components in virtual communities for patients. J Manage Inf Syst 2005;21(4):101-136.
26. 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;7(1):e10 [FREE Full text] [doi: 10.2196/jmir.7.1.e10] [Medline: 15829472]
27. Sutton S. Determinants of health-related behaviors: theoretical and methodological issues. In: Sutton S, Baum A, Johnston M, editors. The Sage Handbook of Health Psychology. London: Sage; 2004:94-126.
28. Festinger LA. A theory of social comparison processes. Hum Relat 1954;7(2):117-140. [doi: 10.1177/001872675400700202]
29. Oudenampsen D, Nederland T, Kromontono E, Pepels R, Sonneveld R, Stukstette M. Patiënten: en consumentenbeweging in beeld. Brancherapport 2006. Utrecht: Nederlandse Patiënten Consumenten Federatie; 2007.
30. Helgeson VS, Taylor SE. Social comparisons and adjustment among cardiac patients. J Appl Soc Psychol 1993 Aug;23(15):1171-1195. [doi: 10.1111/j.1559-1816.1993.tb01027.x]
31. Salzer MS. Consumer-delivered services as a best practice in mental health care delivery and the development of practice guidelines. Psychiatr Rehabil Skills 2002;6(3):355-382. [doi: 10.1080/10973430208408443]
32. Owen JE, Goldstein MS, Lee JH, Breen N, Rowland JH. Use of health-related and cancer-specific support groups among adult cancer survivors. Cancer 2007 Jun 15;109(12):2580-2589 [FREE Full text] [doi: 10.1002/cncr.22719] [Medline: 17503435]
33. van Deursen AJAM, van Dijk JAGM. Measuring digital skills: performance tests of operational, formal, information and strategic Internet skills among the Dutch population. In: Proceedings. 2008 Presented at: 58th Annual Meeting of the International Communication Association; May 22-26, 2008; Montreal, QC, Canada.
34. Preece J, Nonnecke B, Andrews D. The top five reasons for lurking: improving community experiences for everyone. Comput Hum Behav 2004 Mar;20(2):201-223. [doi: 10.1016/j.chb.2003.10.015]
35. Feng J, Lazar J, Preece J. Empathy and online interpersonal trust: a fragile relationship. Behav Inform Technol 2004;23(2):97-106. [doi: 10.1080/01449290310001659240]

Edited by G Eysenbach; submitted 22.12.10; peer-reviewed by B Van den Borne, J Fogel; comments to author 12.04.11; revised version received 01.08.11; accepted 25.08.11; published 07.12.11.

Please cite as:
Van Uden-Kraan CF, Drossaert CHC, Taal E, Smit WM, Bernelot Moens HJ, Van de Laar MAFJ
Determinants of Engagement in Face-to-Face and Online Patient Support Groups
J Med Internet Res 2011;13(4):e106
URL: http://www.jmir.org/2011/4/e106/
doi:10.2196/jmir.1718
PMID:22155649

©Cornelia F Van Uden-Kraan, Constance HC Drossaert, Erik Taal, Willem M Smit, Hein J Bernelot Moens, Mart AFJ Van de Laar. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Benefits of Peer Support in Online Japanese Breast Cancer Communities: Differences Between Lurkers and Posters

Yoko Setoyama[1,2], MS; Yoshihiko Yamazaki[3], PhD; Kazuhiro Namayama[1], PhD

[1]Department of Nursing Informatics, St Luke's College of Nursing, Tokyo, Japan

[2]The Health Care Science Institute, Minato-ku, Tokyo, Japan

[3]Public Health Research Foundation, Institute of Stress Science, Tokyo, Japan

Corresponding Author:
Yoko Setoyama, MS
Department of Nursing Informatics
St Luke's College of Nursing
10-1, Akashi-cho
Chuo-ku,
Tokyo, 104-0044
Japan
Phone: 81 3 3543 6391
Fax: 81 3 5565 1626
Email: yokos-tky@umin.ac.jp

Abstract

Background: Web 2.0 has improved interactions among peers on the Internet, especially for the many online patient communities that have emerged over the past decades. Online communities are said to be particularly beneficial peer support resources for patients with breast cancer. However, most studies of online patient communities have focused on those members who post actively (posters), even though there are many members who participate without posting (lurkers). In addition, little attention has been paid to the usage of online communities among non-English-speaking patients.

Objective: The present study explored the differences in peer support received by lurkers and posters in online breast cancer communities. It also examined the effects of such support on both groups' mental health.

Methods: We conducted an exploratory, descriptive, cross-sectional, Web-based survey among members of four Japanese online breast cancer communities. In an online questionnaire, we asked questions regarding sociodemographics, disease-related characteristics, mental health, participation in online communities, and peer support received from those communities.

Results: Of the 465 people who accessed the questionnaire, 253 completed it. Of the respondents, 113/220 (51.4%) were lurkers. There was no significant difference between lurkers and posters with regard to sociodemographic variables. About half of the posters had been given a diagnosis of breast cancer less than a year previously, which was a significantly shorter period than that of the lurkers (P = .02). The 5 support functions extracted by factor analysis were the same for both posters and lurkers. These were emotional support/helper therapy, emotional expression, conflict, advice, and insight/universality. When the support scores were calculated, insight/universality scored highest for both posters and lurkers, with scores that were not significantly different between the two groups. Among the 5 support scores, emotional support/helper therapy and emotional expression were significantly higher among posters. For posters, emotional support/helper therapy and advice were negatively correlated with the anxiety subscale of the Hospital Anxiety and Depression Scale. Emotional expression, advice, and insight/universality were negatively correlated with the anxiety subscale for lurkers.

Conclusion: We found that posters felt they received more benefits from online communities than lurkers did, including emotional support, helping other patients, and expressing their emotions. Yet even lurkers were found to gain a certain amount of peer support through online communities, especially with regard to advice and insight/universality. The results demonstrate that participation in online communities—even as a lurker—may be beneficial to breast cancer patients' mental health.

(J Med Internet Res 2011;13(4):e122) doi:10.2196/jmir.1696

KEYWORDS

Online support groups; social support; patients; lurkers; breast cancer; mental health; Japan

XSL•FO
RenderX

Introduction

The Internet has become increasingly popular in Japan since the 1990s. The Internet penetration rate in Japan exceeded 75.3% in 2008 [1], and many Japanese people now use the Internet in their daily lives. After the mid 2000s, people began to interact with each other on the Internet using Web 2.0 functions such as blogs, social networking services, and Q&A websites. Web 2.0 is a term that O'Reilly defined as "a set of economic, social, and technology trends that collectively form the basis for the next generation of the Internet, a more mature, distinctive medium characterized by user participation, openness, and network effects" [2]. Users are now able to post comments freely on the Internet without possessing complex technical skills. Hansen [3] stated that Web 2.0 "improved communication and collaboration between people." Specifically, one of the outcomes of the popularization of Web 2.0 was that people with similar health experiences developed online support communities [4]. On the basis of this standpoint, it is important to determine what people do and how they communicate with each other via the Internet over the course of their illness.

Online communities are beneficial because of their availability; for instance, they have no time restrictions [5] and people can access them from any region at no cost [6]. Thus, it is easy for people with disabilities and psychological burdens to receive support from peers online. Previous studies have shown that patients with heart disease [7] and other rare diseases [8] shared their experiences and exchanged emotional and informational support through online communities. In addition, Hill and Weinert [9] found that online communities help participants learn more about themselves, overcome isolation, and find companionship while adapting to their illness. Thus, online communities are now considered a beneficial peer support resource for patients [10].

Since there are many treatment options for breast cancer, patients' informational needs are high. In fact, breast cancer is the most common health topic researched on the Internet. Davison et al [11] reported that support related to breast cancer was the most searched-for health topic on the Internet in the United States, followed by acquired immunodeficiency syndrome, alcoholism, and prostate cancer. Sharf [5] observed that patients with breast cancer exchanged information, social support, and even personal empowerment through online communities. Radin [12] found that breast cancer patients promoted "thick trust" and "collected intelligence" through online communities, and that they discussed various typically painful cancer-related topics with candor, warmth, and even humor [13]. Moreover, online breast cancer communities have been shown to be a useful resource in reducing depression [14,15], dealing with cancer-related trauma [16], and improving posttraumatic growth and psychosocial well-being [14]. Online communities have been found to be comparable in effectiveness with face-to-face support groups [16].

People can participate in online communities in two ways. Those who participate actively are known as posters, and those who do so passively, without making any postings, are known as lurkers [17]. van Uden-Kraan et al [18] found that both posters and lurkers are in some way empowered by participating in online communities; they considered this participation to be a form of bibliotherapy.

Previous researchers have identified some of the reasons why people do not post in online communities, including lack software skills, dislike of the group dynamic, or feeling that the community is a poor fit for them [19]. On the other hand, some people simply do not feel the need to post and feel that they are being helpful by not posting [19]. Many previous studies that describe the benefits of online communities have focused on members who actively contributed by posting messages (ie, posters) [18]. However, Nonnecke and Preece [17] reported that in health-related online communities, an average of 45.5% of people participated as lurkers. If online communities are a peer support resource from which even lurkers can gain some benefit, people who feel that it is a technological and psychological burden to post can use these resources more freely.

To provide further evidence of online communities as a health resource, their effects on users' health should be explored for both posters and lurkers. Moreover, although the Internet penetration rate in Japan is comparable with that of Western countries [20], there have been limited studies of online communities in Japan. Studies of non-English-language online communities are also scarce [21]. Thus, in the present study, we investigated Japanese online breast cancer support communities to determine whether peer support is received differently by lurkers and posters. In addition, we explored the effects of support on members' mental health between the two groups.

Methods

Survey Procedure

In this exploratory, descriptive, cross-sectional study, we conducted a Web survey from September to October 2007, referring to the checklist for the quality improvement of Web surveys [22].

We searched for online communities designed for breast cancer patients using the Google Japan and Yahoo! Japan search engines, which have the largest and second largest numbers of users in Japan, respectively [23]. When searching for online communities, we used the keywords *breast cancer*, *discussion board*, and *mailing list*. Discussion boards and mailing lists are differentiated by registration systems. However, because our research focus was peer support received by members of Internet communities, the registration system distinctions were irrelevant.

We found 12 different breast cancer communities and asked their administrators for survey cooperation via email. During this process, we eliminated those online communities that had participants with non-breast cancers and those in which health care providers served as managers. All of the participating online communities had new posts within 28 days from the start of the survey. Finally, administrators from 4 of the initial 12 online breast cancer communities agreed to cooperate with this survey. The purpose of all of the communities was the exchange of peer support among breast cancer patients.

XSL•FO
RenderX

We developed an online questionnaire form for this open survey. We did not offer any incentive to participate. The four administrators explained the research to their communities' members and provided the questionnaire URL by posting information on their respective community websites. The explanation of the research included a statement about the purpose of this study, the survey duration, and how to store the data on a secure server. We used secure websites to protect personal data. The usability and technical functions of the site were tested by a group of colleagues before we conducted the real test. The 5-page survey site had an average of 8 items on each page of the questionnaire.

Participants were able to navigate to the questionnaire site directly from the community sites by clicking on a hyperlink, and we explained that accessing the questionnaire site would be regarded as an agreement to participate in the survey. To prevent multiple entries from the same individuals, we checked the IP address of everyone who participated in the survey.

Instrument

We did not have a valid instrument to precisely measure social support from peers for posters and lurkers, so we developed a new instrument for the purpose of our study. Of course, there are existing instruments that can be used to measure general social support, such as informational support and emotional support [24], or support networks, such as family members and friends [25]. However, after conducting interviews with seven patients in online breast cancer communities regarding how they used those communities and what kind of support they received from them, we decided to develop a new instrument that could measure social support given specifically by online peers. We then interviewed two nurses in order to check the face validity of our instrument. These nurses were specialists in breast cancer care.

Sociodemographic Characteristics

Our survey inquired about patients' age, marital status (unmarried, married, or separated/widowed), education (middle school, high school, vocational school/2-year college, university/graduate school or higher), and employment (full-time job, housewife, part-time job, or unemployed). All of the participants were women.

Disease-Related Characteristics

The respondents were asked to report on four disease-related characteristics: (1) time since diagnosis of breast cancer (less than 1 year, 1–2 years, 3–5 years, 6–9 years, and 10 years or more, (2) stage of breast cancer at the time of diagnosis (below stage I to beyond stage III), (3) physical symptoms due to breast cancer or breast cancer treatment (eg, pain, feeling tired, arm paralysis, and nausea—respondents who selected more than 1 symptom were categorized as patients with symptoms, and we also counted the total number of symptoms), and (4) personal daily activity level, indicating physical condition. Activity level was indicated using a 5-point Likert scale that ranged from 5, living completely as usual, to 1, almost staying in bed.

Mental Health

Patients rated their levels of anxiety and depression on the Hospital Anxiety and Depression Scale (HADS), which has been used with the general population, cancer patients, and primary care patients [26]. The HADS consists of 14 items: 7 on the depression subscale and 7 on the anxiety subscale. Each item is scored on a 4-point scale from 0 (not present) to 3 (considerable), and the item scores are added, yielding anxiety and depression scores from 0 (minimum symptom load) to 21 (maximum symptom load). A higher score indicates a worse condition. A Japanese version of the scale has been widely used and was confirmed to be reliable and valid [27]. Cronbach alpha for the total HADS score in this study was .89.

Participation in Online Communities

We asked the participants "How often do you post in online communities?" The response items were every time, sometimes, or never—just lurking. We labeled respondents who selected every time and sometimes as posters and those who selected never—just lurking as lurkers.

Received Peer Support

On the basis of our previous interviews, we extracted 8 categories of peer support that study participants received by taking part in online communities. These categories were emotional support, informational support/advice, insight, emotional expression, universality, conflict, empowerment, and helper therapy. Emotional support and informational support were the functions of social support that Cohen et al found in their studies [24]. Insight, universality, modeling, and helper therapy correspond to the concepts that Mishima et al [28], Takahashi et al [29], and Hirose et al [30] found to be the functions of self-help groups. Empowerment corresponds to the study of van Uden-Kraan et al [18]. Conflict has been found to correspond to negative experiences when patients participate in face-to-face support groups [31].

On the basis of these concepts, we formulated 34 items that described the peer support that took place in the online communities. All items had the format of a statement that began with the phrase "Through my participation in online communities..." Respondents could answer on a 5-point Likert scale that ranged from 5 (strongly agree) to 1 (strongly disagree). Emotional support, informational support/advice, insight, and universality were measured with 4 items; conflict was measured with 7 items; empowerment was measured with 4 items; and helper therapy was measured with 3 items.

Analysis

The incidence and average scores of the sociodemographic variables and the current status of participation were calculated for posters and lurkers. Metric variables were analyzed by *t* tests, and categorical variables were analyzed with chi-square tests and Kruskal-Wallis tests. For 61 patients with breast cancer who used online communities, we conducted a pilot test in order to check whether the quantity and quality of the questionnaire was suited to our study's objectives. Then, we revised some of the words to which the patients said they could not respond very well. We also deleted 2 items from insight, 1 item from universality, and 2 items from empowerment because of the

XSL•FO
RenderX

floor and ceiling effects. We thereby used only 29 items to measure peer support received from online communities. We did not include the data of the pilot test samples in the final analysis.

We conducted an exploratory factor analysis to evaluate the factor structure of the support functions for posters and lurkers. While we knew the expected factors based on the previous research used to construct the items, we chose an exploratory factor analysis to determine the best factors for these data. We used principal axis factoring with promax rotation, an oblique rotation method that minimizes the number of variables with high loadings on each factor. This method simplifies the interpretation of the factors. We specified a precedent cut-off of .35 for acceptable factor loadings. To compare the factor constructions between posters and lurkers, we conducted a separate factor analysis for the extracted factors.

After conducting a factor analysis, we deleted 2 items from empowerment, 2 from helper therapy, and 1 from universality because the factor loadings of these items were all less than .35. Considering the factor loadings of each item and the content validity, we extracted 5 factors from the instrument. We then calculated the sum of the scores for each support function, which we referred to as the support score. To compare support scores between posters and lurkers, we conducted an analysis of variance (ANOVA) using a general linear model, controlling for time since diagnosis. We then calculated the Pearson correlation coefficient to determine the relationship between each health status (HADS) and support scores.

Ethical Consideration

We explained the aim of the research project both verbally and in writing to the administrators of the online communities. They were assured that anonymity would be guaranteed and that refusing to participate or withdrawing consent would have no negative consequences. Since the investigation of patients may lead to psychological stress, we made special efforts to reduce the psychological burden of the questionnaire survey and exercised the utmost caution to protect participants' privacy. The Ethics Review Committee of the University of Tokyo approved this study (approval number: 1789).

Results

Participants' Characteristics

The number of visitors to the questionnaire site, or unique site visitors, was 465. We clarified the number of unique visitors based on IP addresses. The number of people who completed the questionnaire was 253. The completion rate, or the ratio of people who agreed to participate to the number of those who finished the survey, was 0.544.

To ensure valid data from a homogeneous sample, we excluded 33 participants: those who had recurrent breast cancer (n = 21), those who had not undergone any surgery for breast cancer (n = 8), and those who had an extremely low daily activity level ("almost staying in bed") (n = 4). Ultimately, we analyzed 220 valid responses. We only analyzed completed questionnaires. The average time in which participants answered the questionnaire was 27 minutes. There were no outliers.

The respondents' active participation in online communities was as follows: every time, n = 14 (6.4%); sometimes, n = 93 (42.2%); and never—just lurking, n = 113 (51.4%).

The characteristics of the survey respondents are shown in Table 1 and Table 2. No variables differed significantly between posters and lurkers. About half of the posters had their breast cancer diagnosis within the previous year, a period that was significantly shorter than that of lurkers (P = .02).

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of posters and lurkers (n = 220) (excluding missing data)

	Posters (n = 107)		Lurkers (n = 113)		*P* value
	n	%	n	%	
Age (years)					.55[a]
≤29	2	2	2	2	
30–39	24	23	30	27	
40–49	60	58	55	50	
50–59	16	15	22	20	
60–69	2	2	2	2	
Mean (SD)	43.71 (7.197)		44.79 (7.474)		.66[b]
Marital status					.24[a]
Unmarried	16	16	30	28	
Married	77	75	62	57	
Separated/widowed	10	10	16	15	
Education					.13[a]
High school	22	21	31	29	
Vocational school/2-year college	34	33	43	40	
University/graduate or higher	47	46	34	31	
Employment					.89[a]
Full-time	30	28	33	30	
Housewife	37	35	32	29	
Part-time	22	21	30	27	
Unemployed	18	17	17	15	

[a] χ^2 test. Degrees of freedom were the number of category –1.

[b] *t* test. Degree of freedom was 219.

XSL•FO
RenderX

Table 2. Health characteristics of posters and lurkers (n = 220) (excluding missing data)

	Posters (n = 107)		Lurkers (n = 113)		*P* value
	n	%	n	%	
Time since di agnosis (years)					.02[a]
<1	52	49	31	38	
1–2	33	31	39	35	
3–5	9	8	23	21	
6–9	10	9	9	8	
≥10	2	2	8	7	
Cancer stage at diagnosis					.39[b]
I	50	47	36	34	
II	43	41	48	45	
III+	8	8	13	12	
Not known	5	5	9	8	
Presence of symptoms [c]					.26[b]
Yes	93	87	85	75	
No	14	13	28	25	
Number of symptoms, mean (SD)	2 (1.685)		2 (1.456)		.62[d]
Physical condition					.77[a]
Living completely as usual	57	53	58	51	
Living as usual	50	47	55	49	
HADS [e], mean (SD)					
Summed scores	12.6 (6.9)		13.4 (8.7)		.52[d]
Depression	6.2 (3.6)		6.5 (4.1)		.63[d]
Anxiety	6.4 (4.1)		6.9 (5.4)		.51[d]

[a] Kruskal-Wallis test.

[b] χ^2 test. Degrees of freedom were the number of category –1.

[c] Respondents checked all of their current symptoms due to breast cancer (eg, pain, tiredness, paralysis of arm, and nausea) and were classified as having symptoms if they chose more than 1 symptom.

[d] *t* test. Degree of freedom was 219.

[e] Hospital Anxiety and Depression Scale.

Support Functions From Online Communities for Posters and Lurkers

The 5 peer support factors that we extracted from the poster and lurker groups were the same (Table 3, Table 4). These 5 factors, which each group felt that they received from peers in their online community, were emotional support/helper therapy, emotional expression, conflict, advice, and insight/universality. Each factor had a Cronbach alpha > .65.

XSL•FO
RenderX

Table 3. Factor analysis of peer support functions for posters (n = 107)

Factor (Cronbach alpha)	Factor loading extracted for each factor
Emotional support/helper therapy (alpha = .752)	
I was encouraged when I was supported by peers	.777
I began to respond positively to my peers	.767
I could talk pleasantly with my peers about topics besides breast cancer	.732
I was encouraged when I could help my peers	.644
I wanted to be as cheerful as my happier peers	.613
I wanted to help other patients who were troubled with breast cancer	.574
I wanted to make others aware of breast cancer	.476
Emotional expression (alpha = .850)	
I could straightforwardly express my feelings about relationships in my workplace or family	.848
I could express my feelings about my relationship with my own doctor	.819
I could straightforwardly talk about my condition	.703
I could express my feelings after breast cancer diagnosis	.518
Advice (alpha = .739)	
I received advice about treatment decision making and the side effects of various treatments	.725
I received advice about day-to-day life with breast cancer, such as a wig and mastectomy bra	.672
I received advice about relationships with family members or colleagues in my workplace	.520
I received advice about my relationship with my doctor and about selecting a hospital	.505
Conflict (alpha = .652)	
I could not express my feelings out of consideration for others	.605
I was concerned that I might get incorrect information about breast cancer	.580
I became tired when breast cancer became the only topic of conversation	.506
I felt discomfort when I was misunderstood by my peers	.497
I regretted that I learned about a better treatment from peers after finishing my treatment	.484
I felt burdened by the time and cost of the peer support resource	.463
I was in trouble when peers recommended I buy some useless products	.383
Insight/universality (alpha = .674)	
I could help myself recover after I realized that my experience was not unique	.688
I had more insight about myself after meeting other patients	.580
I calmed down when I met other patients who had similar experiences to mine	.573

XSL•FO
RenderX

Table 4. Factor analysis of peer support functions for lurkers (n = 113)

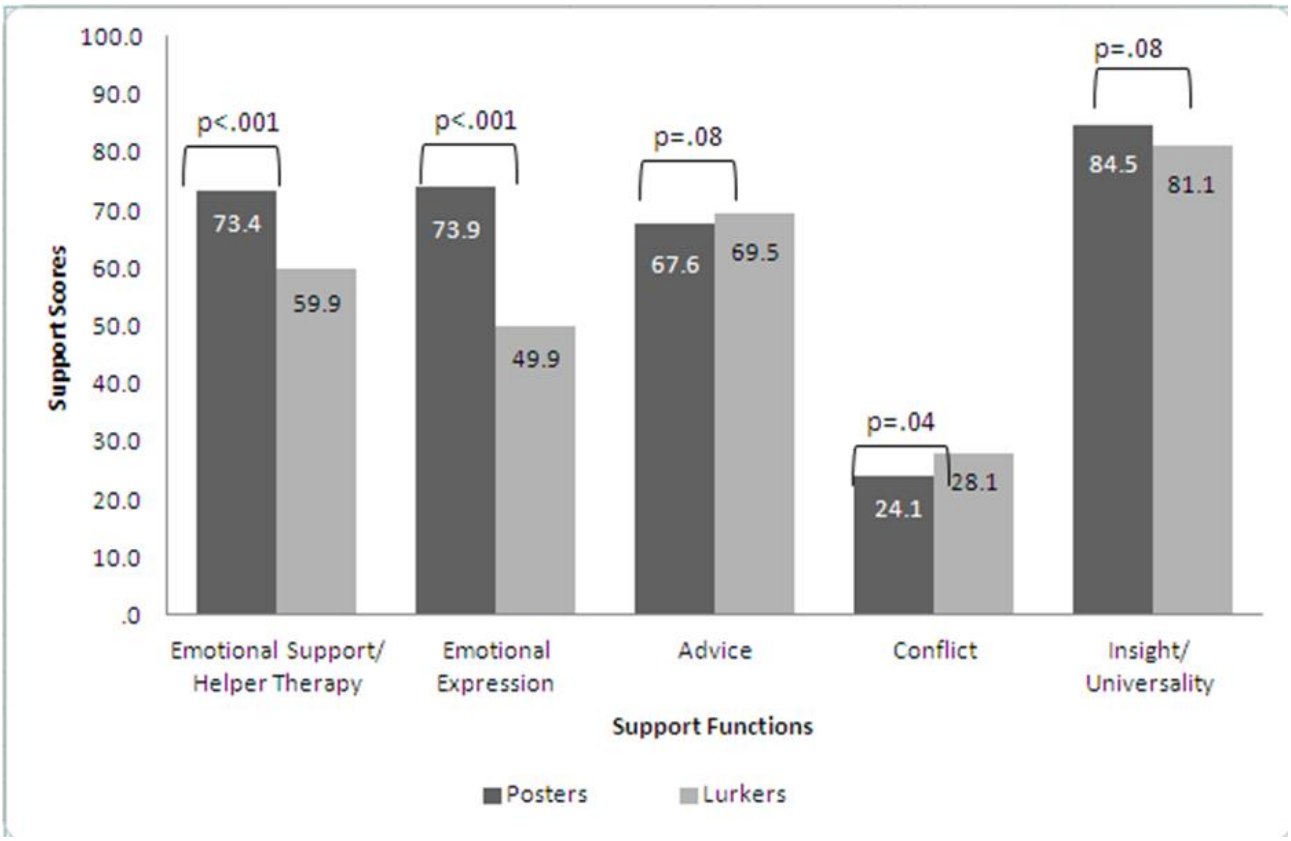
Factor (Cronbach alpha)	Factor loading extracted for each factor
Emotional support/helper therapy (alpha = .786)	
I was encouraged when I was supported by peers	.505
I began to respond positively to my peers	.547
I could talk pleasantly with my peers about topics besides breast cancer	.703
I was encouraged when I could help my peers	.738
I wanted to be as cheerful as my happier peers	.573
I wanted to help other patients who were troubled with breast cancer	.814
I wanted to make others aware of breast cancer	.956
Emotional expression (alpha = .910)	
I could straightforwardly express my feelings about relationships in my workplace or family	.911
I could express my feelings about my relationship with my own doctor	.839
I could straightforwardly talk about my condition	.974
I could express my feelings after breast cancer diagnosis	.925
Advice (alpha = .808)	
I received advice about treatment decision making and the side effects of various treatments	.642
I received advice about day-to-day life with breast cancer, such as a wig and mastectomy bra	.873
I received advice about relationships with family members or colleagues in my workplace	.671
I received advice about my relationship with my doctor and about selecting a hospital	.854
Conflict (alpha = .796)	
I could not express my feelings out of consideration for others	.554
I was concerned that I might get incorrect information about breast cancer	.619
I became tired when breast cancer became the only topic of conversation	.747
I felt discomfort when I was misunderstood by my peers	.767
I regretted that I learned about a better treatment from peers after finishing my treatment	.460
I felt burdened by the time and cost of the peer support resource	.652
I was in trouble when peers recommended I buy some useless products	.735
Insight/universality (alpha = .822)	
I could help myself recover after I realized that my experience was not unique	.926
I had more insight about myself after meeting other patients	.627
I calmed down when I met other patients who had similar experiences to mine	.899

Support Scores of Posters and Lurkers

Each support score, determined based on the extracted factors, is shown in Figure 1. All scores were converted to be out of 100 points. The highest score was for insight/universality for both posters and lurkers. In the results of ANOVA using the general linear model, controlled by time since diagnosis, there was no significant difference between these scores (P = .08). The scores for emotional support/helper therapy (P < .001) and emotional expression (P < .001) were significantly higher for posters.

XSL•FO
RenderX

Figure 1. Support scores for posters and lurkers.



Correlation Between Support Functions and Mental Health

We calculated the correlation between each support and mental health score (HADS) for both posters and lurkers, as shown in Table 5.

For posters, emotional support/helper therapy (r = –.477, P < .001) and advice (r = –.399, P < .001) were negatively correlated with the anxiety subscale. Conflict (r = .287, P = .001) was positively correlated with the depression subscale. For lurkers, emotional expression (r = –.294, P < .001), advice (r = –.655, P < .001), and insight/universality (r = –.495, P < .001) were negatively correlated with the anxiety subscale. Emotional expression (r = –.116, P = .05) also had a slightly negative correlation with the depression subscale.

Table 5. Correlations between support score and mental health as measured by the Hospital Anxiety and Depression Scale (HADS) subscales anxiety and depression (n = 220) (excluding missing data)

	Anxiety		Depression	
	r	*P* value	*r*	*P* value
Posters (n = 107)				
Emotional support/helper therapy	–.477	<.001	.002	.99
Emotional expression	.090	.30	.045	.60
Advice	–.399	<.001	.082	.34
Conflict	.132	.12	.287	.001
Insight/universality	.130	.13	–.007	.93
Lurkers (n = 113)				
Emotional support/helper therapy	.042	.47	.048	.41
Emotional expression	–.294	<.001	–.116	.05
Advice	–.655	<.001	.004	.95
Conflict	.049	.40	.093	.11
Insight/universality	–.495	<.001	–.048	.41

XSL•FO
RenderX

Discussion

Most of the posters who participated in our study had received a breast cancer diagnosis relatively recently. Notably, this result does not match that of the study of van Uden-Kraan et al [18]. The participants in our survey were all patients with breast cancer, which was different from the samples of previous studies that included patients with chronic disease. Patients with breast cancer are faced with major decisions about their treatment during a short period just after their diagnosis. Thus, in the process between diagnosis and decision making, their informational needs are high and they experience psychological distress [32]. The participants in this study may also have had high informational needs before making major decisions or just after beginning their treatment. Thus, it is possible that patients who have a recent diagnosis may use online communities actively as posters in order to ask questions and express their emotions.

Also in contrast to the study of van Uden-Kraan et al [18], in our study we did not find a significant age difference between posters and lurkers. They attributed this difference to the lack of computer skills of older people. However, in Japan, the penetration of the Internet among people 60 years or older has grown rapidly, from 37.6% in 2008 to 58.0% in 2009, so their familiarity with computers has increased [23]. Thus, in Japan, the difference between posters and lurkers is not thought to be influenced by a lack of computer skills resulting from age. Instead, these differences could be influenced by the level of people's informational needs, as mentioned above.

In this study, among the 5 functions of peer support from online communities, emotional support and emotional expression were similar to the peer support provided by face-to-face support groups [24]. These were also defined as group cohesiveness and catharsis in online communities [33]. Goodman [31] defined advice, insight, and universality as peer support functions, while Mishima et al [28] and Vilhauer [33] referred to helper therapy as altruism [33]. Moreover, according to Goodman, conflict is considered to be a negative form of support from peers [31].

To put it simply, we ascertained that the 5 support functions found by this survey characterized social support from peers. Additionally, both posters and lurkers were found to receive some amount of support. Social support plays an important role as a buffer for stressful events such as the diagnosis of a life-threatening disease [34]. Online communities are not just convenient for participants because they are accessible 24 hours a day from anywhere; they also act as a beneficial social support resource, even through passive participation (ie, not posting).

Among the 5 functions, insight/universality scored the highest among both posters and lurkers. Therefore, it can be said that the main function of online communities is to provide insight and universality. In our study, scores for emotional support/helper therapy and emotional expression differed significantly between posters and lurkers. So emotional support/helper therapy and emotional expression may be considered to be support that can be received by actively participating in online communities. However, lurkers received a certain amount of these support functions. It can thereby be said that lurkers can feel comforted by online communities, and that they express their emotions without posting because of the modeling effect. People can identify with others more easily by reading or hearing about experiences that are similar to their own [35]; as van Uden-Kraan et al stated [18], lurking in online communities might be seen as a form of bibliotherapy. In addition, lurkers and posters did not have significantly different scores for advice or insight/universality. These results indicate that lurkers, who participate passively, can receive a similar amount of support to that received by posters through active participation.

In this study, the more posters felt they received emotional support/helper therapy and advice, the less anxious they felt. Furthermore, the more advice lurkers gained from their peers, the less anxious they felt. Learning from others who have had similar experiences helps people control their emotions by reducing the number of future unknowns [36-38]. Because our study was cross-sectional, we cannot explain the causal relationship between them. However, theoretically, social support has a positive influence on people's mental health. Therefore, these associations between received peer support and better mental health may imply that participants reduce their emotional conflict through peer support from online communities. As for advice, people who receive informational support can experience reduced future uncertainty, which can assuage their anxiety. Posters are considered to actively give and receive support, and their actions can positively affect their emotional status. Lurkers can be said to have simulated experiences through reading others' exchanges in posts.

In our study, the more emotional expression lurkers—who do not express their experiences and feelings directly—received, the less anxious they felt. Iwamitsu et al [39] state that expressing negative and positive emotions appropriately could be beneficial for reducing emotional distress among breast cancer patients. Therefore, our study may partially support his opinion. We found associations between more emotional expression and less anxiety only among lurkers because the lurkers probably read the contents of the online community more carefully than the posters did. It may be easier for lurkers to gain more social modeling effects than for posters, who may not read others' posts and only post to meet their own needs. Additionally, Silverberg [40] explained the process of bibliotherapy as knowledge about others' experiences leading to positive outcomes through the mechanism of changes such as insight and catharsis. According to our results, the main function of online communities is to provide insight and universality. Being part of an online community could thereby have a positive effect on mental health. Previously, it was thought that only active participation in online support groups could have a positive effect on mental health [15]. However, this study reveals that online communities may have positive effects for even passive participants as well.

The age group with the most frequent occurrence of breast cancer is women in their 50s [41]. Thus, many patients play multiple roles in their families and careers. It is therefore important to let them know which social support resources can be used with few limitations in terms of time, location, and psychological burden. Moreover, it is important to inform them

XSL•FO
RenderX

that these resources may be beneficial for their mental health, even for passive users.

Limitations

In this study, we asked for cooperation from administrators of online communities found using Google and Yahoo! Japan. Thus, the population of the study sample is considered to contain those who were already Internet users and those who were likely to seek peer support. Additionally, we could not analyze the characteristics of those who did not complete the questionnaire or those who stopped participating in an online community before the samples were recruited. This could mean that people who had a negative impression of online communities eliminated themselves from the survey. Thus, the results may be biased to indicate more positive conditions than those that actually exist. In future, we should identify the characteristics of those who stop using online communities and determine what kind of population is best suited to using this support resource.

Due to the cross-sectional nature of this study, we were unable to determine the causal relationship between received support and mental health. Therefore, it is possible that people with less initial anxiety were more likely to receive peer support. Although it is theoretically reasonable to expect that greater support leads to better health, a longitudinal study is needed to confirm such a causal relationship.

Despite these limitations, this study suggests that even lurkers, who participate passively in online communities, can gain peer support through the Internet, and that some peer support may have a positive effect on their mental health. Health care providers should therefore provide information about online communities as a support resource for patients with breast cancer.

Conflicts of Interest

None declared

References

1. Ministry of Internal Affairs and Communications (Japan). 2009. Communications Usage Trend Survey 2008 (in Japanese) URL: http://www.soumu.go.jp/menu_news/s-news/02tsushin02_000001.html [accessed 2010-10-31] [WebCite Cache ID 5tuD0dAFK]
2. O'Reilly T. O'Reilly Media Inc. 2005. What Is Web 2. 0: Design Patterns and Business Models for the Next Generation of Software URL: http://oreilly.com/web2/archive/what-is-web-20.html [accessed 2010-10-31] [WebCite Cache ID 5tuB03CMj]
3. Hansen MM. Versatile, immersive, creative and dynamic virtual 3-D healthcare learning environments: a review of the literature. J Med Internet Res 2008;10(3):e26 [FREE Full text] [doi: 10.2196/jmir.1051] [Medline: 18762473]
4. Fox S. Online health research 2006. Washington, DC: Pew Internet & American Life Project; 2006 Oct 29. URL: http://www.pewinternet.org/~/media//Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf [accessed 2011-11-30] [WebCite Cache ID 63aWqD5Bg]
5. Sharf BF. Communicating breast cancer on-line: support and empowerment on the Internet. Women Health 1997;26(1):65-84. [doi: 10.1300/J013v26n01_05] [Medline: 9311100]
6. Nápoles-Springer AM, Ortíz C, O'Brien H, Díaz-Méndez M, Pérez-Stable EJ. Use of cancer support groups among Latina breast cancer survivors. J Cancer Surviv 2007 Sep;1(3):193-204. [doi: 10.1007/s11764-007-0029-7] [Medline: 18648970]
7. Bonniface L, Green L. Finding a new kind of knowledge on the HeartNET website. Health Info Libr J 2007 Dec;24 Suppl 1:67-76. [doi: 10.1111/j.1471-1842.2007.00742.x] [Medline: 18005296]
8. Coulson NS, Buchanan H, Aubeeluck A. Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group. Patient Educ Couns 2007 Oct;68(2):173-178. [doi: 10.1016/j.pec.2007.06.002] [Medline: 17629440]
9. Hill WG, Weinert C. An evaluation of an online intervention to provide social support and health education. Comput Inform Nurs 2004 Oct;22(5):282-288. [Medline: 15520598]
10. Hoey LM, Ieropoli SC, White VM, Jefford M. Systematic review of peer-support programs for people with cancer. Patient Educ Couns 2008 Mar;70(3):315-337. [doi: 10.1016/j.pec.2007.11.016] [Medline: 18191527]
11. Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. Am Psychol 2000 Feb;55(2):205-217. [Medline: 10717968]
12. Radin P. “To me, it’s my life”: medical communication, trust, and activism in cyberspace. Soc Sci Med 2006 Feb;62(3):591-601. [doi: 10.1016/j.socscimed.2005.06.022] [Medline: 16039031]
13. 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. [Medline: 12450957]
14. Lieberman MA, Goldstein BA. Self-help on-line: an outcome evaluation of breast cancer bulletin boards. J Health Psychol 2005 Nov;10(6):855-862. [doi: 10.1177/1359105305057319] [Medline: 16176962]
15. Griffiths KM, Calear AL, Banfield M. Systematic review on Internet Support Groups (ISGs) and depression (1): Do ISGs reduce depressive symptoms? J Med Internet Res 2009;11(3):e40 [FREE Full text] [doi: 10.2196/jmir.1270] [Medline: 19793719]

XSL•FO
RenderX

16. Winzelberg AJ, Classen C, Alpers GW, Roberts H, Koopman C, Adams RE, et al. Evaluation of an internet support group for women with primary breast cancer. Cancer 2003 Mar 1;97(5):1164-1173 [FREE Full text] [doi: 10.1002/cncr.11174] [Medline: 12599221]
17. Nonnecke B, Preece J. Lurker demographics: counting the silent. In: Proceedings. 2000 Presented at: SIGCHI Conference on Human Factors in Computing Systems; Apr 1-6, 2000; The Hague, The Netherlands. [doi: 10.1145/332040.332409]
18. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Self-reported differences in empowerment between lurkers and posters in online patient support groups. J Med Internet Res 2008;10(2):e18 [FREE Full text] [doi: 10.2196/jmir.992] [Medline: 18653442]
19. Preece J, Nonnecke B, Andrews D. The top five reasons for lurking: improving community experiences for everyone. Comput Hum Behav 2004 Mar;20(2):201-223. [doi: 10.1016/j.chb.2003.10.015]
20. International Telecommunications Union. 2009 Feb 09. ITU global view URL: http://www.itu.int/ITU-D/connect/gblview/index.html [accessed 2011-12-05] [WebCite Cache ID 63hrIwTHz]
21. Høybye MT, Johansen C, Tjørnhøj-Thomsen T. Online interaction: effects of storytelling in an internet breast cancer support group. Psychooncology 2005 Mar;14(3):211-220. [doi: 10.1002/pon.837] [Medline: 15386774]
22. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
23. Zaidan-houjin internet kyoukai [Foundation of Internet], editors. Internet hakusyo [Internet White Paper] (in Japanese). Tokyo: Kabushiki-gaisya Impress R&B; 2010.
24. Cohen S, Underwood LG, Gottlieb BH. editors. Social Support Measurement and Intervention: A Guide for Health and Social Scientists. Oxford: Oxford University Press; 2000.
25. Iwasa I, Gondo K, Masui S, Inagaki H, Kawai C, Otsuka R, et al. Nihongo ban “social support shakudo” no shinraisei narabini datousei [The validity and reliability of the Japanese version of “Multidimensional Scale of Perceived Social Support”] (in Japanese). Kosei no shihyo 2007;54(6):26-33.
26. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [Medline: 11832252]
27. Zigmond AS, Snaith RP, Kitamura T. HAD shakudo [Hospital Anxiety and Depression Scale] (in Japanese). Seishinka-shindangaku 1993;4:371-372.
28. Mishima I, Kubo H, Noda T. editors. Self-help-group no riron to tenkai [The Theory and Application of Self-help Groups] (in Japanese). Tokyo: Chuou-houki; 2001.
29. Takahashi I, Ogasawara T, Kume Y. Gan taikensya no QOL ni taisuru jijyo group no joutyo teki support no kouka [The effect of self-help group emotional support on cancer survivors' QOL] (in Japanese). Nihon gan kango gakkaishi 2004;18(1):14-24.
30. Hirose H, Hisada M, Aoki Y. Jutsu go nyu gan kanja no tame no tanki support group no kinou ni kansuru shitsu teki kennkyu [A qualitative study on the functions of time-limited support group for post-surgical breast cancer patients: analyses of the group process] (in Japanese). Gan kango 2001;6(5):428-437.
31. Goodman CC. Perceived social support for caregiving: measuring the benefit of self-help/support group participation. J Gerontol Soc Work 1991;16(3-4):163-175. [doi: 10.1300/J083v16n03_13]
32. Suzuki H, Eto Y, Oishi F. Shindan kara shujutsu made no process ni okeru nyugan kanja no shinri henka [Psychological changes in the process of their diagnosis and the operations among patients with breast cancer] (in Japanese). Mie kangogaku zasshi 2008;10:47-57.
33. Vilhauer RP. Perceived benefits of online support groups for women with metastatic breast cancer. Women Health 2009 Aug;49(5):381-404. [doi: 10.1080/03630240903238719] [Medline: 19851944]
34. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]
35. Frieswijk N, Steverink N, Buunk BP, Slaets JP. The effectiveness of a bibliotherapy in increasing the self-management ability of slightly to moderately frail older people. Patient Educ Couns 2006 May;61(2):219-227. [doi: 10.1016/j.pec.2005.03.011] [Medline: 15939567]
36. 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]
37. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
38. van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]
39. Iwamitsu Y, Shimoda K, Abe H, Tani T, Okawa M, Buck R. The relation between negative emotional suppression and emotional distress in breast cancer diagnosis and treatment. Health Commun 2005;18(3):201-215. [doi: 10.1207/s15327027hc1803_1] [Medline: 16187928]

XSL•FO
RenderX

40. Silverberg LI. Bibliotherapy: the therapeutic use of didactic and literary texts in treatment, diagnosis, prevention, and training. J Am Osteopath Assoc 2003 Mar;103(3):131-135 [FREE Full text] [Medline: 12665221]
41. Matsuda T, Marugame T, Kamo K, Katanoda K, Ajiki W, Sobue T, Japan Cancer Surveillance Research Group. Cancer incidence and incidence rates in Japan in 2003: based on data from 13 population-based cancer registries in the Monitoring of Cancer Incidence in Japan (MCIJ) Project. Jpn J Clin Oncol 2009 Dec;39(12):850-858 [FREE Full text] [doi: 10.1093/jjco/hyp106] [Medline: 19797417]

Abbreviations

ANOVA: analysis of variance
HADS: Hospital Anxiety and Depression Scale

Edited by G Eysenbach; submitted 24.11.10; peer-reviewed by J Bender; comments to author 30.01.11; revised version received 12.09.11; accepted 25.09.11; published 29.12.11.

Please cite as:
Setoyama Y, Yamazaki Y, Namayama K
Benefits of Peer Support in Online Japanese Breast Cancer Communities: Differences Between Lurkers and Posters
J Med Internet Res 2011;13(4):e122
URL: http://www.jmir.org/2011/4/e122/
doi:10.2196/jmir.1696
PMID:22204869

©Yoko Setoyama, Yoshihiko Yamazaki, Kazuhiro Namayama. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Website Quality, Expectation, Confirmation, and End User Satisfaction: The Knowledge-Intensive Website of the Korean National Cancer Information Center

Chulmo Koo[1], PhD; Yulia Wati[2], MBA; Keeho Park[3], MD PhD; Min Kyung Lim[3], PhD

[1]College of Business, Chosun University, Gwangju-si, Republic of Korea

[2]Sam M. Walton College of Business, University of Arkansas, Fayetteville, AR, United States

[3]National Cancer Information Center, National Cancer Control Institute, National Cancer Center, Goyang-si, Republic of Korea

Corresponding Author:
Min Kyung Lim, PhD
National Cancer Information Center
National Cancer Control Institute
National Cancer Center
323 Ilsan-ro Ilsandong-gu
Goyang-si, 410-769
Republic of Korea
Phone: 82 31 920 2016
Fax: 82 31 920 2929
Email: mickey@ncc.re.kr

Abstract

Background: The fact that patient satisfaction with primary care clinical practices and physician-patient communications has decreased gradually has brought a new opportunity to the online channel as a supplementary service to provide additional information.

Objective: In this study, our objectives were to examine the process of cognitive knowledge expectation-confirmation from eHealth users and to recommend the attributes of a "knowledge-intensive website.". Knowledge expectation can be defined as users' existing attitudes or beliefs regarding expected levels of knowledge they may gain by accessing the website. Knowledge confirmation is the extent to which user's knowledge expectation of information systems use is realized during actual use. In our hypothesized research model, perceived information quality, presentation and attractiveness as well as knowledge expectation influence knowledge confirmation, which in turn influences perceived usefulness and end user satisfaction, which feeds back to knowledge expectation.

Methods: An empirical study was conducted at the National Cancer Center (NCC), Republic of Korea (South Korea), by evaluating its official website. A user survey was administered containing items to measure subjectively perceived website quality and expectation-confirmation attributes. A study sample of 198 usable responses was used for further analysis. We used the structural equation model to test the proposed research model.

Results: Knowledge expectation exhibited a positive effect on knowledge confirmation (beta = .27, P < .001). The paths from information quality, information presentation, and website attractiveness to knowledge confirmation were also positive and significant (beta = .24, P < .001; beta = .29, P < .001; beta = .18, P < .001, respectively). Moreover, the effect of knowledge confirmation on perceived usefulness was also positively significant (beta = .64, P < .001). Knowledge expectation together with knowledge confirmation and perceived usefulness also significantly affected end user satisfaction (beta = .22 P < .001; beta = .39, P < .001; beta = .25, P < .001, respectively).

Conclusions: Theoretically, this study has (1) identified knowledge-intensive website attributes, (2) enhanced the theoretical foundation of eHealth from the information systems (IS) perspective by adopting the expectation-confirmation theory (ECT), and (3) examined the importance of information and knowledge attributes and explained their impact on user satisfaction. Practically, our empirical results suggest that perceived website quality (ie, information quality, information presentation, and website attractiveness) is a core requirement for knowledge building. In addition, our study has also shown that knowledge confirmation has a greater effect on satisfaction than both knowledge expectation and perceived usefulness.

XSL•FO
RenderX

(J Med Internet Res 2011;13(4):e81) doi:10.2196/jmir.1574

KEYWORDS

eHealth; website quality; end user satisfaction; expectation-confirmation theory; health informatics

Introduction

Background

The Internet is rapidly growing and is increasingly used as an open, anonymous, and democratic source of health information and knowledge [1]. Several studies (eg, [2]) have reported that a large percentage of the population now refers to the Internet to find health-related information as their self-reference [3]. The fact that patient satisfaction with primary care clinical practices and physician-patient communications has decreased gradually has brought a new opportunity to the online channel as a supplementary service to provide additional information [4]. This trend, known as electronic health care (eHealth), has changed the way people search for health-related information. Here, eHealth is defined as the use of the Internet to deliver access to health care information, commerce, clinical care, and other health services [5]. According to Eysenbach [6],

> *eHealth is an emerging field in the intersection of medical informatics, public health,and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a stateofmind, a way of thinking, an attitude, and a commitment for networked, global thinkingto improve health care locally, regionally, and worldwide by using information and communication technology.*

According to the above definition, the knowledge of what consumers find as satisfactory information in the health context has great implications, as customers may act seriously upon this information [3]. On the other hand, observational studies find that Internet users often pay little attention to source credibility when seeking out health information on the Web [1], and a lot of available information is of poor quality [3,7]. Therefore, to ensure that the best and the most accurate, timely, and relevant information is used by consumers, health care organizations have an obligation to attract users to this information [1]. Even though the importance of interactivity in website design is well recognized, studies to understand the benefits of interactive websites to attract customers are rarely developed [8]. In this regard, the information system discipline has been called to develop theories and methods that should prove the usefulness of information in eHealth [9]. Among the limitations of eHealth literature from the information systems (IS) perspective, is that there is no in-depth participatory design research on hospital or health care websites [10], and research is still lacking on the design features and development practices of consumer health information websites [11].

The main purpose of this study was to theorize the attributes of ”knowledge-intensive websites” based on the expectation-confirmation theory (ECT) and integrate these with eHealth from prior IS research. In order to maximize the function of websites as knowledge and information sources, we empirically measure website effectiveness by emphasizing the information as knowledge elements of eHealth services in that the benefits of health care are highly associated with the intrinsic value of information [12]. Secondly, we study patients’ and end users’ online behavior and investigate factors affecting their satisfaction with information provided by health care websites from the information systems perspective. This study adopted ECT to examine the process of cognitive knowledge expectation-confirmation from eHealth users. Even though ECT is widely used in marketing literature to study customer satisfaction (eg, [13]), service marketing (eg, [14]), and information systems (eg, [15-17]), few studies have employed this theory for eHealth services.

To achieve our purposes, we carried out an empirical study at the National Cancer Center (NCC), Republic of Korea (South Korea), by evaluating its official website. By considering the fact that among people with cancer, the Internet has become a major source of health information (eg, [18]), the chosen website is appropriate to our proposed model of a knowledge-intensive website. In addition, as a government funded institution, NCC has a website that may be used to evaluate acceptable criteria of a knowledge-intensive website.

The research reported here also makes several contributions to both research and practice. From a theoretical perspective, we presented the concept of a knowledge-intensive website for eHealth. We proposed and validated a range of criteria needed to establish the knowledge-based website as a main information source for patients and/or Internet users. Second, it extends the ECT in the eHealth context to explain how initial knowledge expectation together with website quality influence knowledge confirmation as an actual knowledge outcome gained by users after assessing the information and how these factors influenced postconsumption expectations that may lead to improved consumer satisfaction, which has not been examined in previous literature. Third, this paper focused on the importance of information and knowledge of an eHealth website, which is a new paradigm in the eHealth research area.

This paper is organized as follows: We begin by presenting the basic concept of expectation confirmation theory. In the third section, we describe our research model and hypotheses development. In the fourth section, we provide a description of the methodology that we relied upon to select and analyze the data, and in the fifth section, we present the results of data analysis. The sixth section presents the discussion of the study’s key findings and its limitations. And in the last section, implications and future research are discussed.

Expectation Confirmation Theory (ECT)

The expectancy confirmation paradigm is primarily cognitive in nature because the comparison process in confirmation judgments requires the deliberate processing of information

XSL•FO
RenderX

[14]. Like the original process of expectation-confirmation in explaining behavior intention, we show that the framework of actual knowledge confirmation begins with individuals' initial expectations of a specific knowledge they may gain prior to the searching process. Thus, individuals accept and use the new knowledge. Following a period of initial consumption, individuals form perceptions about the performance or the website, that is, whether it can improve their knowledge or not. Furthermore, they assess the perceived performance of a website compared with their original expectations and determine the extent to which their expectations are confirmed. Because customers' expectations and perceptions of performance can vary from one to another, confirmation can be positive when actual performance is higher than expectations. In this case the consumer is satisfied. But confirmation can be negative when perceived performance falls short of expectations, and, in this case, the consumer will be dissatisfied [19]. In turn, this level of satisfaction or dissatisfaction will influence intended behavior [15].

Rust et al [20] posit that customer expectations are viewed as distributions, that is, each customer has a probability density function that describes the relative likelihood that a particular quality outcome will be experienced. The satisfaction literature suggests that customers may use different types of expectations when forming opinions about a product's anticipated performance. However, the concept of expectations has raised debate among the scholars. First, the definitions of expectations vary, ranging from the "will expectation" concept to the "should expectation" concept to the "ideal expectation" concept. The *will* expectation concept focuses on forecasting or predicting future performance and is refers to a customer's beliefs of what will happen in the postpurchase period. The *should* expectation concept establishes a normative standard for performance and relates to what a customer believes would happen in the next service encounter. The *ideal* expectation concept is concerned with optimal performance and relates to what a customer wants in an ideal sense [19]. Nevertheless, under ideal expectation, it is theoretically unsound to assume that performance levels that exceed the ideal standard result in higher perceived quality than performance levels that are equal to the ideal standard. Furthermore, if expectation is interpreted to represent a feasible ideal, a positive monotonic linkage between the perception-expectation measure and perceived quality would not be expected when the attributes involved are finite ideal attributes [21]. Second, the concept of expectations ignores the possibility that consumers' expectations change as a consequence of consumers' experience and the impact of changes on subsequent cognitive processes [15]. Third, the concept of satisfaction construct is also ambiguous. Some authors view satisfaction as an attitude (eg, [22]), while others differentiate satisfaction from attitude (eg, [23]).

To tackle these limitations, we measured both preknowledge and postknowledge expectations in one model. While preacceptance expectation is based on secondhand experience (eg, others' opinions or information disseminated through mass media), postacceptance expectation is formed by the customers' firsthand experience and is more realistic [15]. In Bhattacherjee's [15] study, this postacceptance expectation is represented as perceived usefulness. Perceived usefulness can be viewed as individual belief or sum of belief, in that perceived usefulness is a cognitive belief salient to IS use [15,24]. Perceived usefulness is the only belief that consistently influences user intention across temporal stages of information systems use; thus, it is an adequate expectation in the information systems usage context [15]. Perceived usefulness is an important variable affecting users' postadoption decisions since, in this stage, users are likely to reevaluate their early acceptance decisions and make their decisions about continued usage [19].

Methods

Research Model

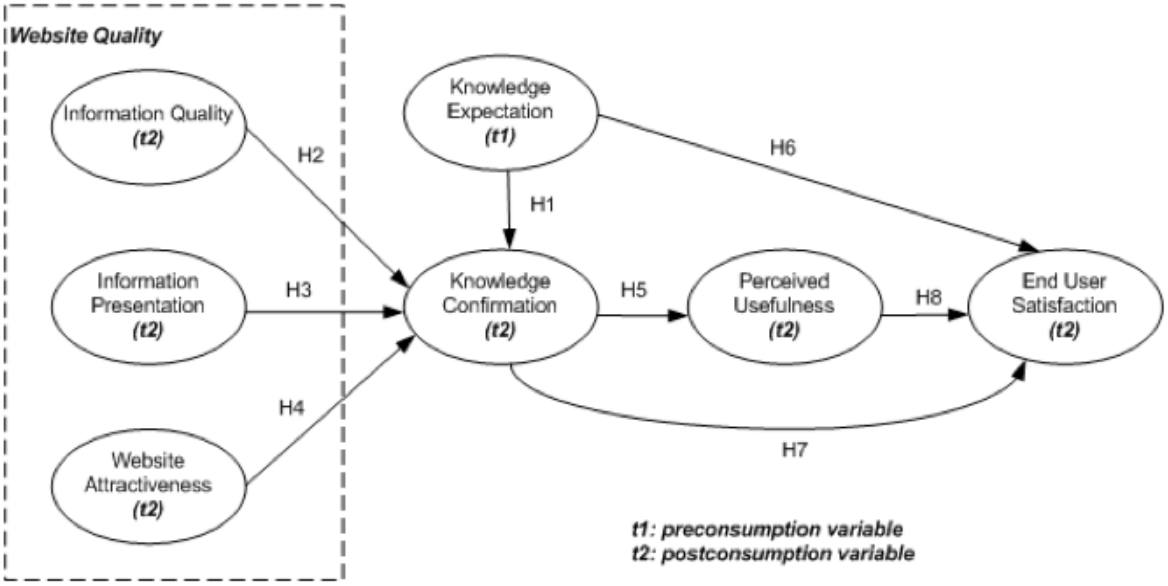
The conceptual model that presents the hypothetical relationships is illustrated in Figure 1. This model shows how knowledge expectations and perceived website quality (ie, information quality, information presentation, and website attractiveness) can influence knowledge confirmation, which leads to perceived usefulness and end user satisfaction. Even though the original concept of expectation confirmation theory strongly suggests perceived performance as an antecedent of satisfaction, we are likely to use *website quality* rather than perceived performance as a predictor of knowledge confirmation. The website's performance in delivering information can be dependent on the quality or nature of the information [25]. To this extent, expectation and confirmation measures focus on personal knowledge or skills, while website quality measures focus on the technical aspects of information. On the basis of ECT discussed previously, we define knowledge expectation as a preconsumption variable (labeled t1 in Figure 1), and the remaining variables as postconsumption variables (labeled t2 in Figure 1). The description of each construct is presented in in Table 1.

XSL•FO
RenderX

Table 1. Definitions of constructs

Construct	Definition
Knowledge expectation (Adapted from [14])	Customers' existing attitudes or beliefs regarding expected levels of knowledge they may gain by accessing the website
Knowledge confirmation (Adapted from [15])	A cognitive belief (the extent to which user's knowledge expectation of information systems use is realized during actual use) derived from prior information systems use
Information quality [26]	Quality of the information system output
Information presentation [27,28]	The degree to which information presentation effectively facilitates interpretation and understanding
Website attractiveness [28,29]	Website's graphic style, that is, the tangible aspect of the online environment that reflects the "look and feel" of the website
Perceived usefulness [24]	An individual's salient belief that using the technology (website) will enhance his or her job performance
Consumer satisfaction [30]	The summary psychological state resulting when the emotion surrounding confirmed or disconfirmed expectations are coupled with the customer's prior feelings about the consumption experience

Figure 1. Research model.



Hypothesis Development

With respect to expectations as comparative referents, it is argued that this expectation influences the confirmation paradigm [31]. In this study, we assumed knowledge expectation to be a user's *will* expectation. Unlike prepurchase behavior conceptualized in marketing concepts, under eHealth services, users have developed positive *will* expectations prior to information access [32]. To support this hypothesis, we have adopted motivation theory (expectancy-value theory) [31] and cognitive consistency theory [33]. First, according to basic motivation theory, motivation may be rooted in the basic need to minimize physical pain and maximize pleasure, and expectancy-value theory suggests that individual expectancies for success and the value they place on achieving it are important determinants of their motivation to perform different achievement tasks [31]. In this case, their positive expectation influences achievement or performance behavior [31]. Second, from a cognitive consistency perspective, the desire of actors to maintain cognitive consistency should affect how they interpret any perceived failures in their counterpart's performance [34]. According to Lord and Maher [33], because actors are more likely to maintain cognitive consistency, those who report high initial expectation will use this as a basis for interpreting the behavior of their counterpart, including the extent to which they determine whether their expectation is met. Similarly, Joyce and Piper [32] have shown that patient expectancy variables are strong predictors of therapy outcomes. Following this prior literature, we also hypothesized that higher knowledge expectation will lead to higher knowledge confirmation (that is, met expectations). Thus, hypothesis 1 is that knowledge expectation is positively associated with knowledge confirmation.

Perceived quality may represent perceived performance of a product or service [20]. Quality assessment relative to expectation represents the most pervasive perspective on quality [35]. From the point of view of expectation, quality is defined by conformance to customer expectation that may relate to excellence, value, and other attributes that are salient to

XSL•FO
RenderX

consumers in shaping their perceptions of quality [35]. The perceived level of quality may either confirm or disconfirm preexpectation. Furthermore, satisfaction is positively affected by expectation and the perceived level of disconfirmation. To this extent, if disconfirmation is perceived to have occurred, then customer satisfaction increases or decreases from the initial expectation [15].

Chiu et al [16] also found that perceived quality is the main predictor of quality confirmation. In our study, we defined three aspects of a website that need to be considered to evaluate knowledge-intensive websites. These are: (1) information quality, (2) information presentation, and (3) website attractiveness. Rogers [36] described information as a difference in matter-energy that effects the uncertainty in situations where a choice exists among a set of alternatives, that is, as matter-energy, information can travel through many forms and channels. In this case, information on the quality of health care is crucial for patients to make informed decisions, and the availability of this information will further empower patients in their relationship with physicians [4]. Information itself contains both extrinsic and intrinsic value that may shape perceptions of quality in the context of use [35]. Following this definition, Nelson et al [35] categorized information quality into four core dimensions: accuracy, completeness, currency, and format. Accuracy refers to the degree to which information is correct, unambiguous, meaningful, believable, and consistent; completeness is the degree to which all possible states relevant to the user population are represented in the stored information; currency represents the degree to which information is up-to-date; and format refers to the degree to which information is presented in a manner that is understandable and interpretable to the user [35].

Bliermel and Hassanein [3] investigated customers' use of the Internet to locate and evaluate health-related information for self-learning, and the result indicated that content quality and technical adequacy played significant roles. Gallant et al [10] investigated the desire content and functionality from the patient-consumer perspective on a hospital website and suggested that website attributes such as visual elements, well-organized personalized information, quality information and reputation, and user-centric design are the important factors to develop eHealth websites. In order to provide a positive user experience, usable technology and the presentation and design of information should be considered as critical factors of website design [10,37]. Jiang and Benbasat [37] examined the effects of various online product presentation formats on consumers' product understanding by specifying two indicators of product understanding performance: consumers' actual product knowledge and perceived website "diagnosticity." The results of this study suggested that the lack of Internet interface to present detailed product information likely leads to customers being less knowledgeable and less informed in making their decision. Thus, we hypothesized the following: (1) hypothesis 2: information quality is positively associated with knowledge confirmation; (2) hypothesis 3: information presentation is positively associated with knowledge confirmation; and (3) hypothesis 3: website attractiveness is positively associated with knowledge confirmation.

Liao et al [19] argued that confirmation during actual use will affect postconsumption expectations such as perceived usefulness. By adopting the concept of cognitive dissonance theory, Bhattacherjee [15] pointed out that users may experience cognitive dissonance or psychological tension if their preacceptance usefulness perceptions are disconfirmed during actual use. Rational users may try to remedy this dissonance by distorting or modifying their usefulness perceptions in order to be more consistent with reality. Thus, confirmation will tend to elevate users' perceived usefulness, and disconfirmation will reduce such perception. Moreover, Jiang and Benbesat [37] posited that the actual knowledge gained by users will positively influence the perceived usefulness of the website. Thus, we hypothesized that knowledge confirmation is positively associated with perceived usefulness (hypothesis 5).

The direct relationship between expectation and customer satisfaction has been proposed in prior research (eg, [14]). According to Bhattacherjee [15], the direct relationship between expectation and satisfaction can be explained by adaptation level theory, which posits that human beings perceive stimuli relative to or as a deviation from an adapted level or baseline stimulus level, where this adapted level is determined by the nature of the stimulus, the psychological characteristics of the individual experiencing that stimulus, and the situational context. The higher the expectation is, the higher one's satisfaction with the service or product, and, conversely, the lower the expectation, the lower one's satisfaction. Thus, we hypothesized that knowledge expectation has a positive effect on end user satisfaction (hypothesis 6).

Confirmation is positively associated with satisfaction as it implies realization of the expected benefits of information systems use, while disconfirmation (to the extent where perceived performance lags expectation) indicates failure to achieve expectation [15]. Through content analysis, Lewis [38] suggested that the use of technology may improve patients' knowledge, involve them in health care decisions, and in turn, lead to better health outcomes. She also posits that the key concern is how to understand the way patients process information and translate it into action. If we can evaluate the best way to deliver the message/information, we will better understand how to use technology to optimize its advantage as a health care learning resource. Major empirical findings also support a positive relationship between expectation and satisfaction (eg, [17,20,30]). Thus, we hypothesized that knowledge confirmation is positively associated with end user satisfaction (hypothesis 7).

Perceived usefulness is the main reason that people decide to use and accept new information systems [10]. Determining the elements of online health information retrieval experience and incorporating those elements in websites that are deemed to contain high quality information from a medical expert's perspective may lead to customer satisfaction [3]. The relationship between perceived usefulness and consumer satisfaction has also been shown by previous studies (eg, [15]). Thus, we proposed that perceived usefulness has a positive effect on end user satisfaction (hypothesis 8).

XSL•FO
RenderX

Measurement of Variables

As far as possible, items used to measure each construct were based on preexisting instruments, and some of these were modified specifically for this study. Information quality items were adopted from Wixom and Todd [39]. We modified items developed by Rai et al [27] and Zhang and von Dran [28] to measure information presentation, and the items for website attractiveness were adapted from Montoya-Weiss et al [29] and Zhang and von Dran [28]. Items for expectation were based on Khalifa and Liu [17]; however, in our study, users were asked to recall the time when they first accessed the website. Moreover, questions for confirmation and end user satisfaction were adopted from Bhattacherjee [15] and Oliver [14], while questions for perceived usefulness were modified from Davis [24] in that respondents were asked to evaluate four forms of information (e-learning, e-book, PowerPoint and multimedia, and testimonial/Q&A format). This research instrument (questionnaire) was checked by academic professors from the information systems department, and a pretest was conducted to ensure the item measures were well communicated and understood. The items used in this study are presented in Multimedia Appendix 1.

Sample and Research Procedure

Our research used the website satisfaction survey, conducted by the National Cancer Center in South Korea. The survey applied a national probability sampling methodology to assess Korean residents' perceptions regarding cancer information and other issues delivered by National Cancer Center. The objective of this survey was to measure customer satisfaction and identify the effectiveness of media usage to distribute the cancer-related information. The questionnaire was administered online by posting the electronic form on the NCC (National Cancer Center) website. When users entered the website, the questionnaire was presented on a new browser window (pop-up window). Data were collected from September 18, 2009 through December 28, 2009. Cash rewards were provided for respondents. Upon the completion of this survey, 200 responses had been collected. In the present study, we excluded data from respondents with an elementary school education level as our *t* test suggested that there was a significantly different perception between this group and the other groups [40], resulting in a study sample of 198 usable responses.

Of the 198 respondents, 71.2% (141) were female. The majority of respondents (100) were from 20 to 29 years of age (50.5%), while 52 were from 30 to 39 years of age (26.3%). More than half (67.7% or 134) of respondents had a university degree. Among the 198 respondents, 47.5% (94) obtained information about cancer information from the Internet, 19.7% (39), from television, 8.6% (17), from family, 8.1% (16) from a medical center, and the remaining 16.2% (32) obtained information about cancer from friends, books, cancer clubs, newsletters, hospital instructions, and other resources. The percentages of respondents that heard about the NCC website by word of mouth and through Internet searches were 44.4% (88) and 40.9% (81) respectively, while others learned of the website from various other sources (eg, brochures, newsletters, advertisements, and recommendations). Furthermore, among the respondents, approximately 59.6% (118) were members of the general population, followed by 22.2% (44) who were family members or other relatives of patients, 14.1% (28) who were researchers/academics, and only 4.0% (8) who were patients. Lastly, we also asked the respondents to indicate how the information they obtained was used. More than 70% (73.2% or 145) of respondents used the information as resource or reference material, while 26.3% (52) and 17.2% (34) used it as self-learning and to educate cancer patients, respectively.

Results

Reliability and Validity

Prior to data analysis, the research instrument was assessed for its reliability as well as its construct validity. Construct validity assessment was performed through confirmatory factor analysis (CFA) using LISREL 8.7 (Scientific Software International, Inc, Lincolnwood, IL). Each scale item was modeled as a reflective indicator of its latent construct. The seven constructs were allowed to covary in the CFA model. First, we checked the scale validity by examining the goodness of fit of the overall CFA model using criteria suggested by Choudhury and Karahanna [41], where the ratio of chi-square to degrees of freedom should not exceed 5; normed fit index (NFI), comparative fit index (CFI), and goodness of fit index (GFI) should be greater than .90; adjusted goodness of fit index (AGFI) should exceed .80; and root mean square error of approximation (RMSEA) should not exceed .80. After exclusion of some invalid items (the third item of information presentation and the first item of knowledge expectation), all indices of goodness of fit (χ^2/df = 1.51; RMSEA = .051; NFI = .91; NNFI = .96; CFI = .97; GFI = .90; AGFI = .83) suggested an adequate model fit of the empirical data. Furthermore, convergent validity was evaluated using three criteria suggested by Fornell and Larcker [42]: (1) all indicator factor loadings should be greater than .70, (2) composite reliabilities (CR) should be greater than .80, and (3) average variance extracted (AVE) should exceed .50. All factor loadings exceeded .70. Composite reliabilities ranged from .84 to .93, and AVE ranged from .63 to .86 (see Table 2). Therefore, all three conditions for convergent validity were met. Lastly, discriminant validity was assessed using criteria recommended by Fornell and Larcker [42], where the square root of AVE should be larger than the correlation scores among constructs. The result indicated that the condition for discriminant validity was also met (see Table 3).

XSL•FO
RenderX

Table 2. Confirmatory factor analysis results

Variable and Item	Item Number	Standardized Solution	Construct Reliability	Average Variance Extracted
Information quality			.92	.68
The website provides accurate information.	IQ1	.84		
The website provides up-to-date information.	IQ2	.74		
The website provides relevant information.	IQ3	.82		
The website provides the content that supports the website's intended purpose.	IQ4	.86		
The website consists of appropriate level of information detail.	IQ5	.88		
Information presentation			.87	.63
The overview, table of contents, and/or summaries/headings are clearly organized.	IP1	.79		
The structure of information presentation is logical.	IP2	.84		
The information presented is understandable.	IP4	.77		
The amount of information presented was just right.	IP5	.78		
Website attractiveness			.93	.69
Overall, the website's color use is attractive.	WA1	.87		
This website has visually attractive screen layouts.	WA2	.87		
This website has an attractive screen background and pattern.	WA3	.85		
This website has eye-catching images or title on homepage.	WA4	.82		
The multimedia contents are attractive.	WA5	.80		
This website is fun to explore.	WA6	.74		
Knowledge expectation			.84	.72
Using this website will increase my knowledge level about cancer-related subjects.	KE2	.87		
Using this website will improve my skills through a learning process.	KE3	.83		
Knowledge c onfirmation			.92	.86
I have learned new knowledge by using this website (as I expected).	KC1	.93		
I have improved my skills by using this website (as I expected).	KC2	.92		
Perceived usefulness			.90	.69
Web tutorial/e-learning	PU1	.89		
Tutorial material in a printable PDF file/e-books	PU2	.85		
PowerPoint slide presentation	PU3	.79		
Testimonial and Q/A content	PU4	.79		
Customer satisfaction			.92	.86
Considering all things, I'm very satisfied with this website.	SF1	.92		
Overall, my interaction with this website is very satisfying.	SF2	.93		

XSL•FO
RenderX

Table 3. Discriminant validity

Variable	IQ	IP	WA	KE	KC	PU	SF
Information quality (IQ)	.83						
Information presentation (IP)	.72	.80					
Website attractiveness (WA)	.51	.65	.83				
Knowledge expectation (KE)	.59	.54	.48	.85			
Knowledge confirmation (KC)	.70	.69	.61	.62	.92		
Perceived usefulness (PU)	.52	.59	.45	.51	.54	.83	
User satisfaction (SF)	.57	.65	.71	.54	.61	.56	.92

We tested the possibility of common method bias by adopting Harman method bias [43]. A single factor explained 40.3% of the variance, indicating no evidence of common method bias. To strengthen our conclusion, we also adopted a marker variable technique recommended by Lindell and Whitney [44]. First, we chose one unrelated criterion variable (user frequency) to serve as the method variance marker variable. Next, we measured the estimation value of the correlation between the method variable and the manifest variables (r_s). Then, we calculated the partial correlation scores (*ryi-M*) using equation 1 (see Textbox 1). We measured the confidence interval using the test statistic (equation 2). Furthermore, using equation 3, we estimated the scores corrected for unreliability and common method variance (CMV) (*r'yi-m*).

Textbox 1. Marker Variable Technique

Equation 1:



Equation 2:



Equation 3:



Where:

$r_{yi\text{-}M}$ is a partial correlation between variable X_i and Y, controlling for M (unmeasured relevant cause),

r_{yi} is the correlation coefficient suspected of being contaminated by the common method variance (CMV),

r_s is the correlation between the method variable and the manifest variable X_1 multiplied by the correlation between the method variable and the manifest variable X_2,

N is the number of samples,

t is the confidence interval,

$r'_{yi\text{-}M}$ is the disattenuated correlation between Y and M,

r_{Yi} is an observed correlation coefficient, and

r_{ii} is the reliability of X_i.

Table 4 presents the correlation results of the measurement constructs. In the first model, we placed perceived usefulness as a dependent variable whereas information quality, information presentation, website attractiveness, knowledge expectation, and knowledge confirmation were predictors. In the second model, perceived usefulness was treated as a predictor of user satisfaction as presented in our original model. As can be seen in Table 4, all relevant predictors have statistically significant correlations with the criterion variable, whereas the theoretically irrelevant predictor has an insignificant correlation with the criterion variable. The correlations of the marker variable (MV) with other predictor variables are low, supporting the discriminant validity of the MV. All the correlation scores remain statistically significant even when CMV is controlled, revealing that these predictors account for theoretically meaningful amounts of the variance. Moreover, the application of equation 3 shows that the disattenuated partial correlations of all four variables with the criterion are slightly higher than the first-order partial correlations, indicating the unreliability always decreases the estimated impact of CMV [44].

XSL•FO
RenderX

Table 4. Hypothetical correlation among constructs (n = 198)

Variable	IQ	IP	WA	KE	KC	UF	PU	SF
IQ	.83							
IP	.72*	.80						
WA	.51*	.65*	.83					
KE	.59*	.54*	.48**	.85				
KC	.70*	.69*	.61*	.62*	.92			
UF	.19	.17	.08	.18	.21	1.00		
PU	.52*	.59*	.45**	.51*	.54*	.24	.83	
SF	.57*	.65*	.71*	.54*	.61*	.24	.56*	.92
r yi-M (PU)	.50*	.57*	.44**	.49**	.52*	.00		
r' yi- M (PU)	.54*	.66*	.47**	.59*	.56*	.00		
r yi-M (SF)	.55*	.64*	.70*	.52*	.59*	.00	.53*	
r' yi- M (SF)	.60*	.73*	.76*	.62*	.64*	.00	.60*	

*$P < .05$
**$P < .01$.

Therefore, we concluded that common method bias does not seem to be a serious problem in this study. Regarding multicollinearity, variance inflation factor (VIF) scores were measured for all constructs as in Gable et al [45]. The VIF scores ranged from 5.57 to 2.30, below the common VIF threshold of 10, indicating all items were subjected to further analysis [45]. Lastly, nonresponse bias was measured by verifying that the early and late respondents were not significantly different [46].

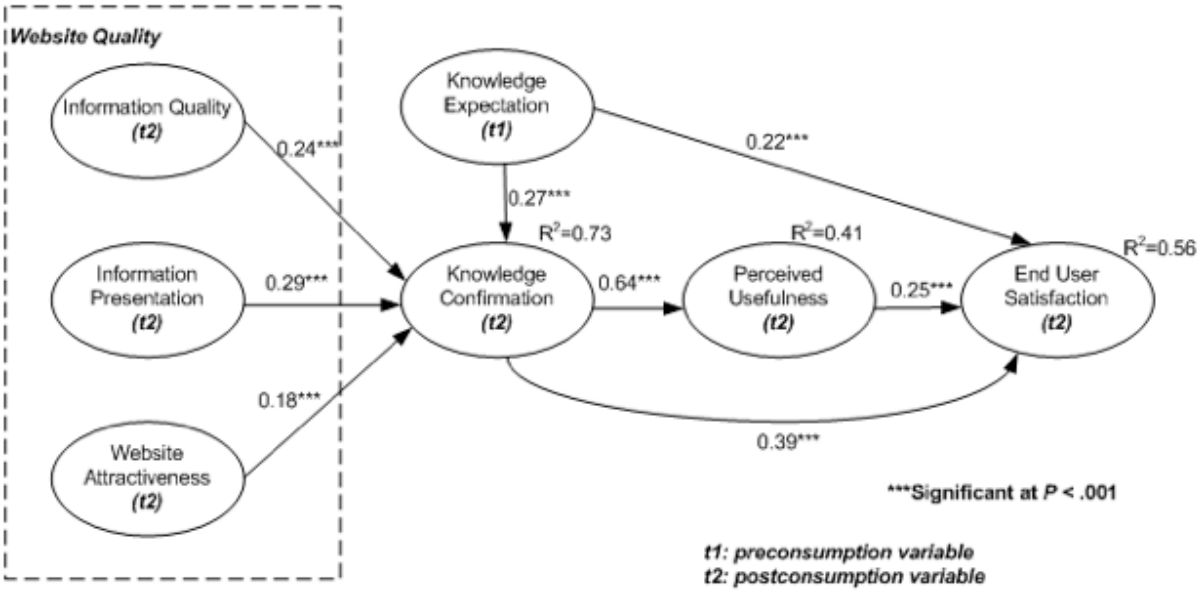
Structural Model and Hypotheses Testing

The structural equation model was used to test the eight hypotheses proposed in this study (see Figure 2). All fit indices have suggested adequate model fit between the proposed model and the actual data (*X2/df* = 1.72; RMSEA = .061; NFI = .90; NNFI = .94; CFI = .95; GFI = .90; AGFI =. 81). As we hypothesized, knowledge expectation exhibited a positive effect on knowledge confirmation (beta = .27, $P< .001$), accepting hypothesis 1. The paths from information quality, information presentation, and website attractiveness to knowledge confirmation were also positive and significant (beta = .24, $P < .001$; beta = .29, $P < .001$; beta = .18, $P < .001$ respectively).

Thus, hypotheses 2, 3 and 4 were accepted. Moreover, the effect of knowledge confirmation on perceived usefulness was also positively significant (beta = .64, $P < .001$); thus, hypothesis 5 was also accepted. Knowledge expectation together with knowledge confirmation and perceived usefulness also significantly affected end user satisfaction (beta = .22, $P < .001$; beta = .39, $P <.001$; beta = .25, $P <.001$ respectively). Hence, hypotheses 6, 7, and 8 were accepted. The model explains 73% of the variance in knowledge confirmation, 41% of the variance in perceived usefulness, and 56% of variance in end user satisfaction.

XSL•FO
RenderX

Figure 2. Hypotheses results.



Discussion

Principal Results and Limitations

Korea has one of the most advanced information technology and IT infrastructure in the world, supporting the diffusion of eHealth technology not only domestically, but also outside the country. Therefore, eHealth has become one of the most important elements for public health care, health informatics, and other related technologies in South Korea. As one of the initial public health care services in this country, the National Cancer Center, initiated by the Ministry of Health and Welfare, South Korea, also delivers its services through the Internet. One of the main functions of this website is providing cancer information in various forms, including electronic learning, e-books, multimedia presentations, and testimonials [47]. As cancer is a common cause of death and its rate is expected to increase in Korea [47], the effort to empirically study the method of delivering health information in Korea may contribute to both research and practice as we proposed in the previous section.

First, our empirical research showed that knowledge expectation was positively associated with actual knowledge confirmed by users after accessing the eHealth website (ie, met expectation). Unlike the traditional ECT in marketing research, our findings confirmed that higher preknowledge expectation may lead to higher postknowledge confirmation. We argued that users or patients' expectations motivate them to access the website, with the assumption that they will gain more knowledge. This finding also supports Joyce and Piper's [32] findings, indicating that initial expectation is a strong predictor of learning outcomes.

Second, this study also found that website quality (ie, information quality, information presentation, and website attractiveness) also influenced the actual knowledge confirmation. Additionally, from our survey, online searching for cancer information is the most popular choice for obtaining information compared with other conventional alternatives. Grounded on this finding, we argued that computer-based information has been an effective strategy for knowledge transfer in the health care context [38]. Moreover, it indicated that as potential patients, website users want to get functional, interactive, and efficient information, that is, knowledge-intensive websites might be the key to enhancing the likelihood of people using health care websites. The website attributes of information presentation and attractiveness are also needed to stimulate learning effectiveness, thus increasing the actual knowledge confirmation [37].

Third, the findings confirmed the positive relationship between knowledge confirmation and perceived usefulness (postexpectation variable) suggesting that users' perceptions of the usefulness of information provided by an eHealth website may be influenced by their confirmation level. Considering the fact that this confirmation level was influenced by website quality, we argued that when the expectation and information quality attributes are both measured in the preconsumption stage, postexpectation is related to information quality [48]. Furthermore, we showed that the usefulness of a website is also be supported by a better design of the website to meet user needs [10].

Fourth, the effects of knowledge expectation, knowledge confirmation, and perceived usefulness on end user satisfaction were also statistically significant. Through these findings, we posit that user satisfaction is determined by expectation of the knowledge and confirmation of expectation following actual use represented by perceived usefulness [15]. Users form this expectation distribution based on their cumulative expectation through postconfirmation, influencing their further perception. To this extent, however, confirmation also showed a greater effect than both preexpectation and postexpectation constructs,

XSL•FO
RenderX

confirming the findings of Bhattacherjee [15] and Oliver [49]. The results also suggested that confirmation may influence satisfaction directly or through the mediation of perceived usefulness, indicating that the relationship between confirmation and satisfaction levels can be modeled in two different ways: using both direct and mediation effects.

Beyond its contribution, this study also has limitations. First, we only investigated the predictor side of satisfaction. Further research is needed to study the outcome side of the satisfaction model (eg, the relationship between satisfaction and intention to use and the relationship between satisfaction and negative word of mouth). Second, even though a range of statistical methods has been used to ensure the validity and reliability of our data, further research is needed to measure the expectation and confirmation at adoption and postadoption to validate the results. Third, this study was based in Korea and used only one specific cancer website. Future research can explore the importance of information and knowledge for different respondents in different countries.

Implications and Future Research

This study provides implications for both research and practice. Theoretical implications of this research are threefold: (1) identification of the attributes of knowledge-intensive websites; (2) enhancement of the theoretical foundation of eHealth from the information systems perspective by adopting ECT; and (3) examination of the importance of information and knowledge and explanation of their impact. First, the raising of concerns about the validity of information on the Internet has been a challenge for eHealth centers whose goal is to provide knowledgeable information presented in an appropriate format and posted on an interactive website. Our study also suggests that an intensive website should be able to influence the cognitive skills of users in learning and absorbing knowledge. Further research may address this initial finding to study how the website attributes presented in this study together with other attributes (eg, service quality) influence consumers' attitudes in a different sense.

Second, this study has enhanced the concept of electronic health care from the information systems perspective by providing theoretical explanations through the adoption of ECT. By demonstrating that preknowledge expectations and perceived information performance influence actual knowledge acquisition, the results indicate that when patients and or users enter a health care website, they bring a certain level of expectation that by accessing and turning on the website, they may improve and gain some new information and knowledge, while explicitly, this process is also influenced by perceived performance. We also argued that during the consumption process, the user's expectations might be adjusted by confirmation, resulting in greater or lower postexpectation beliefs (perceived usefulness). Thus, our study suggests the important linkage of these variables for eHealth satisfaction literature. We measured preexpectation with *will* expectation and the confirmation results showed the *met* expectation condition. However, by considering the ambiguity of the original expectation concept, future research should examine this theory in greater depth.

Third, this study examined the online information performance construct by studying its effects on influencing the knowledge cognition process. Recognizing that transfer of knowledge to patients or end users may help them to participate in the decision-making process toward their health condition, we suggest that further research is needed to examine the roles of other information media, such as mobile information services. Moreover, it is also a challenge for information systems researchers to become involved actively in this area, particularly to examine how to deliver health information in various electronic formats.

Practically, our empirical results indicate that information performance is a core requirement for knowledge building. Through this study, we argued that having accurate, high quality cancer or general health care information published on a reliable website can provide individuals with knowledge and help the consumers to obtain more useful materials. Furthermore, this research suggests that information on eHealth websites should be presented attractively. Online health care can also provide an opportunity for health care centers to learn how to provide online information innovatively to attract more patients or Internet users. The information presentation in various formats (eg, multimedia/power point and e-book) can utilize multiple sensory channels to convey information to users, which, in turn, builds respective mental representations in both verbal and nonverbal system [50]. Thus, eHealth providers should consider these attributes to build their knowledge-intensive websites. Another implication is related to consumer satisfaction. Our study pointed out that confirmation has a greater effect on satisfaction than other variables. Thus, it is not sufficient for a health care center to just improve its quality information attributes, it must also try to meet patients' expectation, and in turn, increase their actual confirmation.

Acknowledgments

The authors gratefully acknowledge the financial support of the National Cancer Center, Korea (Grant numbers NCC-0960350 and NCC-1010131), and the Ministry of Health and Welfare, Republic of Korea.

Conflicts of Interest

None declared

XSL•FO
RenderX

Authors' Contributions

Min Kyung Lim originated the study and supervised all aspects of its implementation. Chulmo Koo had responsibility for the design, analysis, and write-up for this study. Yulia Wati assisted with analyses and writing the article. Keeho Park oversaw the collection and analysis of claims data and gave critical input into the design of the study and interpretation of the results.

Multimedia Appendix 1

Measurement Items.

[PDF File (Adobe PDF File), 90 KB - jmir_v13i4e81_app1.pdf]

References

1. Bates BR, Romina S, Ahmed R, Hopson D. The effect of source credibility on consumers' perceptions of the quality of health information on the Internet. Med Inform Internet Med 2006 Mar;31(1):45-52. [doi: 10.1080/14639230600552601] [Medline: 16754366]
2. Sciamanna CN, Clark MA, Houston TK, Diaz JA. Unmet needs of primary care patients in using the Internet for health-related activities. J Med Internet Res 2002 Dec;4(3):E19 [FREE Full text] [doi: 10.2196/jmir.4.3.e19] [Medline: 12554550]
3. Bliemel M, Hassanein K. Consumer Satisfaction with Online Health Information Retrieval: A Model and Empirical Study. e-Service Journal 2006;5(2):53-83.
4. Altinkemer K, Prabuddha D, Zafer O. Information system and health care XII: toward a customer-to-healthcare provider (C2H) electronic marketplace. Communications of the Association for Information Systems 2006;18(1):413-430.
5. Maheu MM, Whitten P, Allen A. E-Health, Telehealth, and Telemedicine: A Guide to Start-up and Success. San Francisco, CA: Jossey-Bass; 2001.
6. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
7. 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]
8. Jang Z, Chan J, Tan B, Chua W. Effects of interactivity on website involvement and purchase intention. Journal of the association for information system 2010;11(1):34-59.
9. Chang I-C, Hwang HG, Hung MC, Lin MH, Yen DC. Factors affecting the adoption of electronic signature: Executives' perspective of hospital information department. Decision Support Systems 2007;44(1):350-359. [doi: 10.1016/j.dss.2007.04.006]
10. Gallant L, Cynthia I, Kreps G. User-centric hospital websites: A case for trust and personalization. e-Service Journal 2006;5(2):5-26.
11. Ketchum AM. Consumer health information Websites: a survey of design elements found in sites developed in academic environments. J Med Libr Assoc 2005 Oct;93(4):496-499. [Medline: 16239947]
12. Fitterer R, Rohner P, Mettler T, Winter R. A taxonomy for multi-perspective ex ante evaluation of the value of complementary health information systems-applying the unified theory of acceptance and use of technology. Washington, DC: IEEE Computer Society; 2010 Presented at: 43rd Hawaii International Conference on System Sciences; Jan 5-8, 2010; Kauai, HI.
13. Anderson EW. Cross-category variation in customer service and retention. Marketing Letters 1994;5(1):19-30.
14. Oliver RL. A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of Marketing Research 1980;17(4):460-469.
15. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Quarterly 2001;25(3):351-370.
16. Chiu C-M, Hsu M-H, Sun S-Y, Lin T-C, Sun P-C. Usability, quality, value and e-learning continuance decisions. Computers & Education 2005;45(4):399-416. [doi: 10.1016/j.compedu.2004.06.001]
17. Khalifa M, Vanissa L. Determinants of satisfaction at different adoption stages of Internet-based services. Journal of the Association for Information Systems 2003;4(5):206-232.
18. Huang GJ, Penson DF. Internet health resources and the cancer patient. Cancer Invest 2008 Mar;26(2):202-207. [doi: 10.1080/07357900701566197] [Medline: 18259953]
19. Liao C, Chen J-L, Yen DC. Theory of planning behavior (TPB) and customer satisfaction in the continued use of e-service: An integrated model. Computers in Human Behavior 2007;23(6):2804-2822. [doi: 10.1016/j.chb.2006.05.006]
20. Rust RT, Inman JJ, Jia J, Zahorik A. What you don't know about customer-perceived quality: the role of customer expectation distributions. Marketing Science 1999;18(1):77-92. [doi: 10.1287/mksc.18.1.77]
21. Teas K. Expectations, performance evaluation, and customers' perception of quality. The Journal of Marketing 1993;57(4):18-34.
22. LaTour S, Peat N. Conceptual and methodological issues in consumer satisfaction research. Advances in Consumer Research 1979;6:431-437.

XSL•FO
RenderX

23. Hunt H. CS/D-overview and future research directions. In: Hunt H, editor. Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction. Cambridge, MA: Marketing Science Institute; 1977:455-488.
24. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-339.
25. McKinney V, Yoon K, Zahedi FM. The measurement of Web-customer satisfaction: an expectation and disconfirmation approach. Information Systems Research 2002;13(3):296-315. [doi: 10.1287/isre.13.3.296.76]
26. DeLone W, McLean E. The DeLone and McLean model of information systems success: A ten-year update. Journal of Management Information Systems 2003;19(4):9-30.
27. Rai A, Lang SS, Welker RB. Assessing the validity of IS success models: An empirical test and theoretical analysis. Information Systems Research 2002;13(1):50-69. [doi: 10.1287/isre.13.1.50.96]
28. Zhang P, von Dran GM. Satisfiers and dissatisfiers: A two-factor model for website design and evaluation. Journal of the American Society for Information Science 2000;51(14):1253-1268. [doi: 10.1002/1097-4571(2000)9999:9999<::AID-ASI1039>3.0.CO;2-O]
29. Montoya-Weiss MM, Voss GB, Grewal D. Determinants of online channel use and overall satisfaction with a relational, multichannel service provider. Journal of the Academy of Marketing Science 2003;31(4):448-458. [doi: 10.1177/0092070303254408]
30. Oliver RL, Linda G. Effect of satisfaction and its antecedents on consumer preference and intention. Advances in Consumer Research 1981;8(1):88-93.
31. Wigfield A. Expectancy-value theory of achievement motivation: A developmental perspective. Educational Psychology Review 1994;6(1):49-78.
32. Joyce AS, Piper WE. Expectancy, the therapeutic alliance, and treatment outcome in short-term individual psychotherapy. J Psychother Pract Res 1998;7(3):236-248 [FREE Full text] [Medline: 9631345]
33. Lord R, Maher K. Leadership and Information Processing: Linking Perceptions and Performance. London, England: Routledge; 1993.
34. Adobor H. Trust as sensemaking: the microdynamics of trust in interfirm alliances. Journal of Business Research 2005;58(3):330-337. [doi: 10.1016/S0148-2963(03)00077-8]
35. Nelson R, Todd P, Wixom B. Antecedents of information and system quality: an empirical examination within the context of data warehousing. Journal of Management Information System 2005;21(4):199-236.
36. Rogers E. Diffusion of Innovations. New York, NY: Free Press; 1995.
37. Jiang Z, Benbasat I. The effects of presentation formats and task complexity on online consumers' product understanding. MIS Quarterly 2007;31(3):475-500.
38. Lewis D. Computer-based approaches to patient education: a review of the literature. J Am Med Inform Assoc 1999;6(4):272-282. [Medline: 10428001]
39. Wixom BH, Todd PA. A theoretical integration of user satisfaction and technology acceptance. Information Systems Research 2005;16(1):85-102. [doi: 10.1287/isre.1050.0042]
40. Wilson E, Balkan S, Lankton N. Current trends in patients' adoption of advanced e-health services. Washington, DC: IEEE Computer Society; 2010 Presented at: 43rd Hawaii International Conference on System Sciences; Jan 5-8, 2010; Kauai, HI.
41. Choudhury V, Karahanna E. The relative advantage of electronic channel: A multidimensional view. MIS Quarterly 2008;32(1):179-200.
42. Fornell C, Larcker D. Evaluating structural equations with unobservable variables and measurement error. Journal of Marketing Research 1981;18(1):39-50.
43. Harman H. Modern Factor Analysis. Chicago, IL: University of Chicago Press; 1976.
44. Lindell MK, Whitney DJ. Accounting for common method variance in cross-sectional research designs. J Appl Psychol 2001 Feb;86(1):114-121. [Medline: 11302223]
45. Gable G, Sedera D, Chan T. Enterprise systems success: a measurement model. In: Proceedings of the International Conference on Information Systems. Atlanta, GA: Association for Information Systems; 2003 Presented at: 24th International Conference on Information Systems; 2003; Seattle, WA.
46. Armstrong J, Overton T. Estimating nonresponse bias in mail surveys. Journal of Marketing Research 1977;14:396-402.
47. National Cancer Control Institutue, National Cancer Center. Cancer Facts and Figures 2009 in the Republic of Korea. Goyang, Republic of Korea: National Cancer Center and Ministry of Health and Welfare, Republic of Korea; 2009:226-231.
48. Oliver RL, Burke RR. Expectation processes in satisfaction formation. Journal of Service Research 1999;1(3):196-214. [doi: 10.1177/109467059913002]
49. Oliver R. Cognitive, affective, and attribute bases of the satisfaction response. Journal of Customer Research 1993;20(3):418-430.
50. Jiang Z, Benbasat I. The effects of presentation formats and task complexity on online consumers' product understanding. MIS Quarterly 2007;31(3):475-500.

XSL•FO
RenderX

Abbreviations

AGFI: adjusted goodness of fit index
AVE: average variance extracted
CFA: confirmatory factor analysis
CFI: comparative fit index
CMV: common method variance
CR: composite reliability
df: degrees of freedom
ECT: expectation confirmation theory
GFI: goodness of fit index
IP: information presentation
IQ: information quality
IS: information systems
KC: knowledge confirmation
KE: knowledge expectation
MV: marker variable
NCC: National Cancer Center
NFI: normed fit index
PU: perceived usefulness
RMSEA: root mean square error of approximation
SF: satisfaction
UF: user frequency
VIF: variance inflation factor

Edited by G Eysenbach; submitted 20.05.10; peer-reviewed by A Bhattacherjee, A Vézina, N Zhang; comments to author 13.11.10; revised version received 07.12.10; accepted 25.03.11; published 02.11.11.

Please cite as:
Koo C, Wati Y, Park K, Lim MK
Website Quality, Expectation, Confirmation, and End User Satisfaction: The Knowledge-Intensive Website of the Korean National Cancer Information Center
J Med Internet Res 2011;13(4):e81
URL: http://www.jmir.org/2011/4/e81/
doi:10.2196/jmir.1574
PMID:22047810

©Chulmo Koo, Yulia Wati, Keeho Park, Min Kyung Lim. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 02.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Communications Between Volunteers and Health Researchers during Recruitment and Informed Consent: Qualitative Content Analysis of Email Interactions

Anne Townsend[1], MA, PhD; Zubin Amarsi[1], BSC(Hons); Catherine L Backman[1,2], PhD; Susan M Cox[3], MA, PhD; Linda C Li[1,4], PhD

[1]Arthritis Research Centre of Canada, Vancouver, BC, Canada

[2]Occupational Science and Occupational Therapy, University of British Columbia, Vancouver, BC, Canada

[3]Maurice Young Centre for Applied Ethics, University of British Columbia, Vancouver, BC, Canada

[4]Department of Physical Therapy, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Anne Townsend, MA, PhD
Arthritis Research Centre of Canada
895 10th Avenue West
Vancouver, BC, V5Z 1L7
Canada
Phone: 1 604 871 4564
Fax: 1 604 879 3791
Email: atownsend@arthritisresearch.ca

Abstract

Background: While use of the Internet is increasingly widespread in research, little is known about the role of routine electronic mail (email) correspondence during recruitment and early volunteer–researcher interactions. To gain insight into the standpoint of volunteers we analyzed email communications in an early rheumatoid arthritis qualitative interview study.

Objectives: The objectives of our study were (1) to understand the perspectives and motivations of individuals who volunteered for an interview study about the experiences of early rheumatoid arthritis, and (2) to investigate the role of emails in volunteer–researcher interactions during recruitment.

Methods: Between December 2007 and December 2008 we recruited 38 individuals with early rheumatoid arthritis through rheumatologist and family physician offices, arthritis Internet sites, and the Arthritis Research Centre of Canada for a (face-to-face) qualitative interview study. Interested individuals were invited to contact us via email or telephone. In this paper, we report on email communications from 12 of 29 volunteers who used email as their primary communication mode.

Results: Emails offered insights into the perspective of study volunteers. They provided evidence prospectively about recruitment and informed consent in the context of early rheumatoid arthritis. First, some individuals anticipated that participating would have mutual benefits, for themselves and the research, suggesting a reciprocal quality to volunteering. Second, volunteering for the study was strongly motivated by a need to access health services and was both a help-seeking and self-managing strategy. Third, volunteers expressed ambivalence around participation, such as how far participating would benefit them, versus more general benefits for research. Fourth, practical difficulties of negotiating symptom impact, medical appointments, and research tasks were revealed. We also reflect on how emails documented volunteer–researcher interactions, illustrating typically undocumented researcher work during recruitment.

Conclusions: Emails can be key forms of data. They provide richly contextual prospective records of an underresearched dimension of the research process: routine volunteer–researcher interactions during recruitment. Emails record the context of volunteering, and the motivations and priorities of volunteers. They also highlight the "invisible work" of research workers during what are typically considered to be standard administrative tasks. Further research is needed to fully understand the role of routine emails, what they may reveal about volunteers' decisions to participate, and their implications for research relationships—for example, whether they have the potential to foster rapport, trust, and understanding between volunteer and researcher, and ultimately shift the power dynamic of the volunteer–researcher relationship.

(J Med Internet Res 2011;13(4):e84) doi:10.2196/jmir.1752

XSL•FO
RenderX

KEYWORDS

Emails; recruitment; informed consent; volunteer-researcher interactions; rheumatoid arthritis; qualitative research; motivations to volunteer; help-seeking; self-management

Introduction

The Internet is extending research designs [1] and transforming data generation techniques [2] revealing ethical and methodological issues around recruitment [3,4]. A core tool in this transformative process is email, which is routinely used in research [5]. Emails leave an audit trail of volunteer–researcher (or research workers such as research assistants/coordinators) interactions, and record priorities and concerns of volunteers [6]. They also offer insight into the context of a volunteer's daily life and illness experience, therefore providing a prospective record (rather than a retrospective account) of why or how volunteers decide to participate (or not) in health research.

Email communications may influence volunteer–researcher relationships. Kvale [7] suggests that email interviews potentially alter the power imbalance of participant–researcher interactions, by offering opportunities for a more respectful and symmetrical relationship due to "a shared context of communication" (researcher and participant share a flow of information in their own time and space). This potential shift in the dynamics of research relations may extend to email communications during the recruitment and consent process. For example, email correspondence may extend beyond straightforward recruitment information exchange. Volunteers may disclose negative illness/help-seeking experiences and see volunteering as a way of accessing otherwise unavailable medical advice. Role boundaries can be blurred in these circumstances, and researchers need to be sensitive to the context of participation, but be clear that they cannot offer therapy or medical advice. Also, as emails contain sensitive and personal details, secure and confidential storage is important. All of these factors highlight the skill requirements needed by research workers (eg, research coordinators and assistants) whose role is often seen as purely administrative [8].

Recently, Internet recruitment and email interviews have been the focus of research that has identified ethical, methodological, and practical issues [9]. The potential of routine email correspondence in health research, however, remains unexplored. We addressed this gap by focusing on the content and context of email communications to explore their role in the recruitment and informed consent process and their implications for the volunteer–researcher relationship.

Methods

Study Design and Participants

This research formed part of a larger qualitative interview study: Early Rheumatoid Arthritis Help-Seeking Experience (ERAHSE). Rheumatoid arthritis is a chronic musculoskeletal condition and a major cause of disability. The main symptoms are pain, stiffness, joint swelling, and fatigue characterized by exacerbations and remissions [10]. If treatment is delayed, damage can occur in other organs, including the heart and lungs. Onset can be sudden or gradual, and the focus of care is to control the symptoms and limit disease and debility. Timely treatment is crucial to avoid irreversible joint damage, which may lead to permanent disability, increased personal suffering, and medical cost [11].

Individuals who had a diagnosis of rheumatoid arthritis in the previous 12 months, lived in the province of British Columbia (BC), Canada, and were English speakers were eligible. Volunteers were recruited, via information leaflets, through local arthritis clinics and websites of the provincial arthritis organizations. Recruitment documents were also sent to eligible persons by their family practitioner or rheumatologist's office. The information leaflet invited people to share their experiences of early rheumatoid arthritis in an interview study.

Data Collection

The information leaflet provided phone numbers and email addresses for 2 members of the research team who would conduct interviews. Out of a sample of 38 participants, 29 used email as their main form of communication. Email was used to confirm eligibility, provide consent forms, discuss queries, and schedule interviews. Typically email communications continued over a period of 1 or 2 weeks. However, in some cases email correspondence extended to several weeks or months due to practical difficulties of scheduling, illness, or life events. The emails were password protected and stored securely on the server of the Arthritis Research Centre of Canada. Factual information (eg, how participants heard about the research) and general comments (eg, length and tone of emails and broad content) were recorded in field notes. Due to the number and content of emails, we recognized they were a rich source of data. We subsequently sought permission to analyze the correspondence of 15 participants, who had engaged in the most email contact overall. The analysis presented here is based on the email communications of the 12 who provided consent (11 females, 1 male). Ethical approval was received from the University of British Columbia Behavioral Research Ethics Board. Volunteers were invited to choose their own pseudonyms, which have been used.

Data Analysis

The analysis was iterative. A thematic approach informed by a constant comparative method was used. AT and ZA read and coded the emails independently. After discussion and repeated readings of the emails, three initial themes were identified, compared with field notes, and examined for consistency across data types (email, interviews, and field notes). Focusing on emails, further discussion led to agreement on higher-order themes. Constant comparison across data types and scrutiny of all data independently prior to team discussion added rigor to the analysis, contributing to validity of the data-driven claims.

XSL•FO
RenderX

Results

Volunteer emails varied in number, content, length, and style. Most volunteers noted their diagnostic status and willingness to take part. Some indicated an interest in participating in the ERAHSE study but requested more information about research tasks and the potential risks and benefits of participation. The majority elaborated on their illness situation beyond the eligibility criteria. Several gave richly contextual accounts of their symptoms, medication use, interactions with health professionals, and navigating the health care system [12-15]. Here we focus on four main themes arising from the data: (1) research participation as reciprocity, (2) volunteering as self-managing and help-seeking, (3) ambivalence around participation and informed consent, and (4) practical considerations of participation.

Research Participation as Reciprocity: Mutual Benefits

Volunteers expressed mutual benefits of participating. In describing their experiences and contributing to the knowledge base, they wanted to help the research initiative and others with early rheumatoid arthritis. At the same time they hoped to benefit from sharing their stories and securing advice or information about illness management. The email below illustrates the perceived twin benefits of participation, for the volunteer and the research endeavor. Nicole volunteers to help the research in the face of frustrating symptoms. She anticipates that sharing her story might help herself and the research, and hopes to gain insight into disease and pain management:

> *Hello Anne*
>
> *My name is Nicole. I am a 33 year-old woman, who was diagnosed with rheumatoid arthritis in about September of last year. I have just, in the last few weeks made contact with the Arthritis Society and received the emailed newsletter, in which I read about the information you are gathering from newly diagnosed patients. I would be happy to talk with someone about my experiences if it would be helpful. I suggest that, as I have been frustrated lately with the disease and with managing pain, that it would help me to talk about it and hopefully gain some insight that I have been missing.* [Nicole]

During the subsequent interviews several individuals elaborated on their email disclosures. They expressed hope that their experiences could help our research and future rheumatoid arthritis patients, and assist medical professionals in offering care. At the same time they hoped to benefit through gaining advice and information about available resources. In contrast, some participants reflected on how they made contact solely as a help-seeking strategy, in the face of frustrated attempts to access timely care for worsening symptoms.

Volunteering as Self-Managing and Help-Seeking

In their initial emails, some volunteers expressed helplessness about their symptoms and frustration at formal health care, making no mention of our study. Some had gained a diagnosis from their family physician and were waiting to see a rheumatologist for effective medications. Several had sought information online and recognized the need for, but were unable to gain, prompt treatment. These volunteers had been induced to contact us in the face of unpredictable, severe, debilitating, or abnormal symptoms and rising anxiety about their situation. The email below illustrates uncertainties around symptoms and concerns about obtaining a timely meeting with a rheumatologist:

> *Regarding my arthritis: a few weeks ago I got inflammation and swelling in both my thumbs. Then 4 weeks ago my finger next to my thumb swelled like a cigar and has stayed that way. Then the joint swelled and became sore and I can see after only a month my finger twisting. Within the last month I have pain in both shoulders as well and in the bone by my wrist as well as my left small toe. IS THIS NORMAL TO COME ON SO FAST? I asked my Dr to send me to a rheumatologist and he told me there was a one-year waiting list to see one. I am in tears and very sad to see my finger twisting right in front of my eyes and I cannot get to see a specialist.* [Nicolette]

Nicolette described (at interview) the context in which she had emailed us. She had suspected she was in the early stages of rheumatoid arthritis, due to previous knowledge about the disease and an Internet search, which identified the importance of a prompt diagnosis and early treatment. Given this knowledge, and being told by her family physician that there would be a delay of 12 months prior to seeing a rheumatologist, she felt frustrated and sought further information on the Internet. She then found our study and contacted us to talk “to someone” and gain advice.

As illustrated above, some individuals were prompted to volunteer for our study due to frustrated attempts at formal help-seeking. They viewed research participation as a way of accessing much needed support and advice in their quest for prompt treatment. Emails raised questions such as “What should I do?” and included comments such as “I might learn something [if I take part]”, indicating volunteer need (for support in and access to help-seeking) and perceived benefits of participation. This posed potential ethical problems for free and informed consent.

Ambivalence Around Participation

Email correspondence revealed questions about informed consent and offered some insight into the decision-making process. Rain made contact hoping for help in navigating the health care system and was in “two minds” about participating. In his initial email, he described his frustration with the health care he was receiving and did not refer to the research study. Subsequently, he asked a series of questions about accessing care while considering whether to take part. Rain emphasized that his primary motivation for making contact was his hope for a speedy rheumatologist referral, which was difficult to obtain in his rural community. His ambivalence about participating was apparent in his questions about the informed consent document, as he asked explicitly about the risks and benefits:

> *Hi Anne*

XSL•FO
RenderX

> *Thank you for your concern, I was originally looking for a study for a cure or therapy...*
>
> *I would like to get some understanding about the section – Risks and Potential Benefits–in the consent form it mentions "It is possible that some topics discussed may raise new and sometimes difficult issues..." What sort of issues should one be concerned about? As mentioned above, I'm looking for a cure to get rid of my daily pain. An interview may help you, but I'm still suffering.*
>
> *Thank you.* [Rain]

Rain elaborated in his interview that he "went on the Internet to look for some support" [15, p 23]. The email record offers a glimpse of the prospective decision-making process from the volunteer perspective, rather than through hypothetical or retrospective concerns around consent—for example, when eliciting responses during an interview. This is a vivid illustration that informed consent is a 2-way flow that extends beyond ensuring volunteers have received the consent form prior to interview, to review at the time of interview. It also highlights the need to be flexible regarding communication formats. In this example the volunteer agreed to a phone conversation regarding his concerns and any potential risks, burdens, or benefits of participating for him, compared with potential benefits for others and the research more generally.

Practical Considerations of Participation

After receiving the recruitment documents, several volunteers focused on practical aspects of participation. They reported busy lives characterized by symptom management and hospital appointments, and gave insight into the research experience as they negotiated a convenient time and place for interview. The emails also offered volunteers the opportunity to set out the parameters of participation. In the correspondence below, Teresa notes her preferred location and three suitable times for the interview, asking the researcher to let her know what "works best":

> *Hello,*
>
> *I looked over the attachments and everything's ok. After consideration I think it would be too long a day for me to add the interview into a (hospital) appointment. The [occupational therapist] appointments tend to go on for 1 ½ hours or more. My rheumatoid arthritis is very active right now and I'm easily fatigued. I would prefer a home interview, which would be more relaxed and give better insight into the impact of my rheumatoid arthritis. Possible dates are Monday April 7 1pm, Tuesday April 8 10 am or Friday April 11 10 am.... Let me know what works best.* [Teresa]

This email records a daily life, compromised by symptoms and treatment, and adds context to the data generated at interview. It records the potential burdens and costs of the research task for this participant in real time (ie, the efforts taken in order to participate in a research interview study). It also highlights the importance of a convenient time and place for collecting data, for the comfort of the participant and the quality of the data gained.

Discussion

Principal Results

The email communications offered insight into the perspectives of volunteers in our study. They generated prospective data on motivations to take part, recruitment, and informed consent. First, we found there was a reciprocal element to participating. Some volunteers felt they could be of help to research and at the same time hoped that participating would be of help to them. Second, others were prompted to volunteer due to their acute need for information in the face of troublesome symptoms and frustrations with the health care system. For these individuals volunteering was solely a self-management or help-seeking strategy. Third, some ambivalence was illustrated when deciding to participate, regarding the difference between potential benefits for the volunteer and benefits for the research in general. Fourth, practical difficulties of participation arose—for example, scheduling an interview in the context of a daily life organized around symptom containment and medical appointments. Finally, the emails also revealed rarely discussed dimensions of the volunteer–researcher interactions and the invisible work of researchers. Overall, our findings contribute new knowledge to the scant information on the ethics involved in email communications [16,17].

Context of Volunteering: Experience of Early Illness and Help-Seeking

In common with those in other studies, our volunteers hoped to gain health benefits [18]. People who are in the early stages of a chronic illness may well experience uncertainty and anxiety about their condition [12,14,15]. Such feelings may be exacerbated when people are not provided with a firm diagnosis or prompt treatment [12]. In this context, people who feel that they are unheard in the health care system and are aware that they require timely treatment may be inclined to volunteer for research about their condition. Our recruitment materials described an interview study and an interest in personal experiences. This may well suggest an outlet for a personal illness story to be heard and promise hope of advice or support in a patient's quest for a speedy diagnosis or effective treatment. More research is needed to identify how far people volunteer for research to access information or advice as part of their self-management and help-seeking strategies.

Recruitment, Informed Consent, and Volunteer–Researcher Interactions

The emails provided prospective records of aspects of recruitment, consent, and volunteer–researcher interactions, and as such generated data on an underinvestigated dimension of the research process. Emails facilitate a 2-way flow of information exchange, in a "shared context of communication" [7], and have the potential to contribute to a more collaborative health research relationship in the era of the informed epatient [19]. The volunteers in our study had the opportunity to interact in their own time and space at their own convenience, rather than in a face-to-face situation or via the more immediate and

XSL•FO
RenderX

(possibly) intrusive telephone. This may have shifted the balance of control and offered the possibility of an active volunteer [6] participating in a more meaningful and involved recruitment and informed consent process. For example, individuals may have been more inclined to enter into a prolonged dialogue about participation and to broach "sensitive" issues when deciding whether to take part. The volunteer emails in our research recorded reasons for taking part (Nicole) and doubts, reflections, and questions about benefits and risk (Rain). Volunteers may also share emotional stories and frustrations (eg, Nicolette) or, on a practical basis, take the initiative in terms of when and where the interview should take place (eg, Teresa). In future research projects it would be instructive to ask the participants their views on the role of email communications with researchers and research workers during recruitment.

Potential for Rapport Building Through Email Communications

The opportunities for building more collaborative and dynamic relationships in electronic health care [20] applies equally to the qualitative health research process. "The emergent nature of many qualitative studies makes the achievement of rapport with participants and feelings of interpersonal trust crucial to the generation of questions considered important or interesting by both parties" [21 p3]. For this community of volunteers with newly diagnosed rheumatoid arthritis, emails offered a way to engage in dialogue at an anxious and frustrating time, and provided opportunities to foster trust and rapport (eg, one participant requested to be interviewed by the researcher with whom she had been emailing) [14]. This potential benefit mirrors what Childress [22] describes as "proximal benefit:" both the participant and researcher can potentially gain from building an appropriate but respectful relationship during the research process.

The Invisible Work of Research Workers

The volume and content of emails surprised us, motivated this analysis, and highlighted the undocumented or invisible work of researchers and research coordinators [8]. Given the context of some volunteers' illness stories and navigating the health care system, such communications may create expectations regarding the nature of the response. Although we needed to be sensitive to volunteers' circumstances, and emails offered opportunities to build trust and rapport, we also needed to negotiate the boundaries between acting as "therapist" and acting as "sensitive researcher." This necessitated time for prompt, careful, and informative correspondence as we attempted to achieve a careful balance between (objective) pragmatism and (subjective) empathy in the context of multiple tasks and deadlines. Also, given the context of some volunteers' illness stories and their reported frustrations at "not being heard," research workers need perhaps to consider how volunteers are informed that they do not fulfill study criteria.

Practical Obstacles to Participation

The emails pinpointed the practical costs of participation for the volunteers. The real-time communications suggested that for these participants, the burdens of taking part in interviews held some practical obstacles—for example, the potential to aggravate symptoms such as fatigue. More research is needed to assess whether such considerations are relevant beyond this dataset and for volunteer patients in all types of health research.

Limitations

Our findings are limited in scope. As in all qualitative research we do not claim to make generalizations, but to gain a more in-depth understanding of social phenomena. We drew on a small number of emails from study volunteers with newly diagnosed rheumatoid arthritis, who perhaps were particularly keen to participate due to their help-seeking and illness experiences. We highlight that this analysis of the email communication included 1 male and 11 females and could not undertake a gender comparison. In the future, it would be preferable to study a gender-balanced sample. It would also be beneficial to include all of the emails from volunteers rather than a selection. We suggest, however, that the findings can be usefully explored in a range of research settings and designs. Below we offer some observations regarding practice, future research, and educational initiatives.

Practice

Standard operating procedures created at study inception are one way to ensure all research staff approach communications in a thoughtful and consistent way. These procedures should include mechanisms for secure handling and storage of emails, and it should remain clear that study participants be informed of the risks to privacy when using email and that they may prefer (and should be offered) alternative means of communication.

Suggestions for Research

To better understand the role of email use in health research, we need more evidence on emails and how this form of communication may influence recruitment, informed consent, and volunteer–researcher relationships, as well as the skill set needed by research workers. Exploratory research questions could include "What is the nature and extent of routine email communications in different research populations and what are the potential challenges and benefits?"

Educational Aspects

On the basis of our findings, we cannot recommend extensive educational interventions. However, we suggest that educational workshops, which focus on emerging ethical issues in the use of new technologies, could include sessions on email communications. Workshops could engage those involved in research, such as ethicists, health research participants, and researchers, to identify and reflect on emerging issues. Comparisons could be made between phone and email communications in terms of ease of use, content, form and language, and interpersonal relations. Topics could include issues around using emails as data and securing consent to do so.

Conclusion

We are unaware of other studies that have been based on the analysis of volunteer recruitment emails. The emails tracked part of the decision-making process in real time, recording volunteers' hopes, concerns, and practical contingencies in the context of their illness experience. Because emails can be a rich,

XSL•FO
RenderX

prospective data source, researchers may wish to include them as data, which has implications for consent. Although this is a small sample, from which we cannot make general statements, volunteers in other contexts may see health research participation as a way to access care, information, and advice. Research workers should be aware of this during the recruitment and informed consent process.

Emails are not only a technological development but also a reformulation of recruitment and informed consent offering the potential for increased dialogue during routine communications in health research. A key implication of this study is how email communications revealed the invisible work of research workers during recruitment and informed consent. Using the emails as data improved our understanding of the decision-making process, the context in which people volunteered for our study, and the practical obstacles involved. There was a suggestion that emails fostered opportunities for meaningful and thoughtful dialogue over time, but more research is needed to investigate this and perhaps their capacity to shift the dynamics from a traditional to a more symmetrical relationship, as well as a more considered informed consent process.

Acknowledgments

We would like to thank Canadian Institutes of Health Research (CIHR) for funding this research. We also thank Katie Adam at the Arthritis Research Centre of Canada, Vancouver, for her support in preparing this manuscript, and Professor Michael McDonald at the Maurice Young Centre for Applied Ethics, University of British Columbia, and Joan Sieber, Emerita Professor of Psychology at California State University, for their valuable and insightful comments on earlier drafts of the paper. We also thank Paul Adam for his help in data collection and contributions to discussions. We express our gratitude to the volunteers who participated in this research.

Conflicts of Interest

None declared

Authors' Contributions

AT had the original idea, and led data collection, analysis, interpretation and writing. AT guided ZA in the early drafts and subsequently wrote all later drafts of the paper.
ZA wrote the first draft of the paper and worked with AT on subsequent early drafts. ZA participated in data collection, analysis, interpretation, and writing.
CB commented on all drafts of the paper and contributed to conceptual development and structure.
SC commented on all drafts of the paper.
LL was Principal Investigator of the ERAHSE project, commented on all drafts of the paper and contributed to conceptual development and structure.

References

1. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002 Nov;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
2. 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 [FREE Full text] [doi: 10.1136/bmj.328.7439.564] [Medline: 15001506]
3. Kitchin HA. Research Ethics and the Internet: Negotiating Canada's Tri-Council Policy Statement. Black Point, NS: Fernwood; 2007.
4. 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] [doi: 10.2196/jmir.7.1.e6] [Medline: 15829478]
5. Cooper C. The use of electronic mail in the research process. Nurse Res 2000;7:24-30.
6. Rhodes SD, Bowie DA, Hergenrather KC. Collecting behavioural data using the world wide web: considerations for researchers. J Epidemiol Community Health 2003 Jan;57(1):68-73 [FREE Full text] [Medline: 12490652]
7. Kvale S. InterViews: An Introduction to Qualitative Research Interviewing. Thousand Oaks, CA: Sage; 1996.
8. Davis AM, Hull SC, Grady C, Wilfond BS, Henderson GE. The invisible hand in clinical research: the study coordinator's critical role in human subjects protection. J Law Med Ethics 2002;30(3):411-419. [Medline: 12497701]
9. Hamilton RJ, Bowers BJ. Internet recruitment and e-mail interviews in qualitative studies. Qual Health Res 2006 Jul;16(6):821-835. [doi: 10.1177/1049732306287599] [Medline: 16760538]
10. Newman S, Fitzpatrick R, Revenson TA, Skevington S, Williams G. Understanding Rheumatoid Arthritis. London: Routledge; 1996.
11. Kumar K, Daley E, Carruthers DM, Situnayake D, Gordon C, Grindulis K, et al. Delay in presentation to primary care physicians is the main reason why patients with rheumatoid arthritis are seen late by rheumatologists. Rheumatology (Oxford) 2007 Sep;46(9):1438-1440 [FREE Full text] [doi: 10.1093/rheumatology/kem130] [Medline: 17578850]

XSL•FO
RenderX

12. Townsend A, Adam P, Cox SM, Li LC. Everyday ethics and help-seeking in early rheumatoid arthritis. Chronic Illn 2010 Sep;6(3):171-182. [doi: 10.1177/1742395309351963] [Medline: 20610465]
13. Li LC, Townsend A, Adam PM, Cox SM, Amarsi Z, Backamn CL, the ERAHSE Team. Crossing the threshold: help-seeking for early symptoms in people with rheumatoid arthritis. (Abstract 2027). In: Arthritis Rheum. 2009 Presented at: The Annual Conference of American College of Rheumatology (ACR); November, 2009; Philadelphia.
14. Townsend A, Cox SM, Li LC. Qualitative research ethics: enhancing evidence-based practice in physical therapy. Phys Ther 2010 Apr;90(4):615-628 [FREE Full text] [doi: 10.2522/ptj.20080388] [Medline: 20203093]
15. Townsend A. Working to manage chronic illness in daily life. Occup Ther Now. 2011 Sep. (5) p. 20-22 URL: http://www.caot.ca/otnow/sept%2011/working.pdf [WebCite Cache ID 62PBii78Y]
16. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans 1998 (with 2000 and 2005 amendments). Ottawa: Public Works and Government Services Canada; 2002. URL: http://www.pre.ethics.gc.ca/archives/tcps-eptc/docs/TCPS%20October%202005_E.pdf [WebCite Cache ID 62PA98bvZ]
17. Kralik D, Warren J, Price K, Koch T, Pignone G. The ethics of research using electronic mail discussion groups. J Adv Nurs 2005 Dec;52(5):537-545. [doi: 10.1111/j.1365-2648.2005.03612.x] [Medline: 16268860]
18. McDonald M, Townsend A, Cox SM, Paterson ND, Lafrenière D. Trust in health research relationships: accounts of human subjects. J Empir Res Hum Res Ethics 2008 Dec;3(4):35-47. [doi: 10.1525/jer.2008.3.4.35] [Medline: 19385755]
19. Ferguson T, Frydman G. The first generation of e-patients. BMJ 2004 May 15;328(7449):1148-1149 [FREE Full text] [doi: 10.1136/bmj.328.7449.1148] [Medline: 15142894]
20. Kaplan B, Litewka S. Ethical challenges of telemedicine and telehealth. Camb Q Healthc Ethics 2008;17(4):401-416. [doi: 10.1017/S0963180108080535] [Medline: 18724880]
21. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. Chapter 10: Qualitative research. Ottawa, ON: CIHR, NSERC, SSHRC; 2010. p. 135-146 URL: http://www.pre.ethics.gc.ca/pdf/eng/tcps2/TCPS_2_FINAL_Web.pdf [WebCite Cache ID 62PBEEH3j]
22. Childress H. The anthropologist and the crayons: changing our focus from avoiding harm to doing good. J Empir Res Hum Res Ethics 2006 Jun;1(2):79-88. [doi: 10.1525/jer.2006.1.2.79] [Medline: 19385880]

Abbreviations

ERAHSE: Early Rheumatoid Arthritis Help-Seeking Experience

Edited by G Eysenbach; submitted 07.02.11; peer-reviewed by M Alam, M Graber, S Sethi, M Spriggs; comments to author 12.04.11; revised version received 22.06.11; accepted 24.06.11; published 13.10.11.

Please cite as:
Townsend A, Amarsi Z, Backman CL, Cox SM, Li LC
Communications Between Volunteers and Health Researchers during Recruitment and Informed Consent: Qualitative Content Analysis of Email Interactions
J Med Internet Res 2011;13(4):e84
URL: http://www.jmir.org/2011/4/e84/
doi:10.2196/jmir.1752
PMID:21997713

©Anne Townsend, Zubin Amarsi, Catherine Backman, Susan M Cox, Linda C Li. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 13.10.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Ethical Principles for Physician Rating Sites

Daniel Strech[1], MD, PhD

Institute for History, Ethics and Philosophy of Medicine, CELLS - Centre for Ethics and Law in the Life Sciences, Hannover Medical School, Hannover, Germany

Corresponding Author:
Daniel Strech, MD, PhD
Institute for History, Ethics and Philosophy of Medicine
CELLS - Centre for Ethics and Law in the Life Sciences
Hannover Medical School
Carl-Neuberg Str. 1
Hannover, 30625
Germany
Phone: 49 5115322907
Fax: 49 5115325650
Email: strech.daniel@mh-hannover.de

Abstract

During the last 5 years, an ethical debate has emerged, often in public media, about the potential positive and negative effects of physician rating sites and whether physician rating sites created by insurance companies or government agencies are ethical in their current states. Due to the lack of direct evidence of physician rating sites' effects on physicians' performance, patient outcomes, or the public's trust in health care, most contributions refer to normative arguments, hypothetical effects, or indirect evidence. This paper aims, first, to structure the ethical debate about the basic concept of physician rating sites: allowing patients to rate, comment, and discuss physicians' performance, online and visible to everyone. Thus, it provides a more thorough and transparent starting point for further discussion and decision making on physician rating sites: what should physicians and health policy decision makers take into account when discussing the basic concept of physician rating sites and its possible implications on the physician–patient relationship? Second, it discusses where and how the preexisting evidence from the partly related field of public reporting of physician performance can serve as an indicator for specific needs of evaluative research in the field of physician rating sites. This paper defines the ethical principles of patient welfare, patient autonomy, physician welfare, and social justice in the context of physician rating sites. It also outlines basic conditions for a fair decision-making process concerning the implementation and regulation of physician rating sites, namely, transparency, justification, participation, minimization of conflicts of interest, and openness for revision. Besides other issues described in this paper, one trade-off presents a special challenge and will play an important role when deciding about more- or less-restrictive physician rating sites regulations: the potential psychological and financial harms for physicians that can result from physician rating sites need to be contained without limiting the potential benefits for patients with respect to health, health literacy, and equity.

(J Med Internet Res 2011;13(4):e113) doi:10.2196/jmir.1899

KEYWORDS

Physician rating sites; ethics; patient empowerment; patient-physician relationship

Introduction

Physician rating sites allow patients to evaluate their experience and satisfaction with their health care providers, similar to other service-oriented businesses. The ratings are posted online and are intended as a source of information for people searching for a physician.

In addition to the more than 30 private physician rating sites [1,2], more and more publicly hosted physician rating sites have gone online in the last 5 years. In 2007, the National Health Service (NHS) in the United Kingdom launched the NHS Choices website, which allows patients to evaluate both physicians and hospitals. In 2010, the largest German health insurer (AOK) launched its own portal, Arzt-Navi, which initially went through a test phase in 3 out of 16 German states and has been open to all German residents since May 2011. In the United States, the Hospital Compare site, maintained by the Centers for Medicare & Medicaid Services (CMS) and other publicly funded sites, provides information on the quality of care, but it does not yet permit patients to rate physicians [3].

XSL•FO
RenderX

Increasingly, research results are being published on fundamental characteristics of physician rating sites in their present condition in regard to their frequency, content, and user assessment patterns [1,2,4-7]. However, direct evidence of potential benefits and harms of physician rating sites is still lacking.

Only a handful of discussion papers on this topic have been published in scientific journals [3,8-10]. However, the media are increasingly discussing opportunities and challenges of physician rating sites. On the one hand, government and insurance company representatives often express their support of such concepts [11]. Ben Bradshaw, the former British Minister for Health, for example, criticized the general lack of transparency as an argument supporting physician rating sites, stating that "I would never think of going on holiday without cross-referencing at least two guide books and using Trip Advisor. We need to do something similar for the modern generation in healthcare." Other critics have referred to evidence related to questions similar to those of physician rating sites [8]. They highlight that key clinical measures and outcomes are closely linked to patient satisfaction [12,13] and that systematic feedback changes doctors' clinical performance [14]. Nevertheless, physician representatives tend to argue against physician rating sites. Laurence Buckman, Chairman of the British Medical Association's General Practitioners Committee, fears that physician rating sites could compromise physicians: "A website on which people can slander or praise irresponsibly is the wrong approach" [11]. Likewise, Frank Ulrich Montgomery, President of the German Medical Association, described these websites as "platforms for denunciation" [15].

Taking the current state of discussion and scientific analysis of physician rating sites into account, health policy decision making, with respect to the implementation and regulation of physician rating sites, is challenging for at least two major reasons: (1) the lack of outcomes research in the field of physician rating sites, and (2) the controversial but poorly structured (ethical) debate on the pros and cons of physician rating sites.

This paper has two aims. First, it aims to structure the ethical debate around the basic concept of physician rating sites—that is, allowing patients to rate, comment, and discuss physicians' performance, online and visible to everyone. This provides a more thorough and transparent (and therefore more reasonable) starting point for further discussion and further decision making on physician rating sites: what should physicians and health policy decision makers take into account when discussing the basic idea of physician rating sites and its possible implications for the physician–patient relationship? Second, it discusses where and how the preexisting evidence from the partly related field of public reporting of physician performance can serve (at least) as an indicator for specific needs of evaluative research in the field of physician rating sites.

While this paper focuses on the preceding ethical discussion concerning the basic concept of physician rating sites, it does not analyze the wide range of issues surrounding the safety and validity of information provided by physician rating sites. It also does not analyze the various judicial aspects of physician rating sites such as ownership or liability, for example. Needless to say, controversy in eHealth ethics often relates to these implementation issues [16] and they are also relevant for physician rating sites. The issues related to the implementation and management of physician rating sites, however, become relevant and need to be discussed thoroughly after the basic concept of physician rating sites has been generally accepted by physicians and health policy decision makers (as is the case with the basic concept of online patient information about diseases and treatment options, for example).

The following analysis is based on three ethical principles that are presented in widely acknowledged ethical frameworks for health care and health policy (eg, [17,18]). The two major reasons for choosing these rather general ethical principles are the following: (1) the discussion on physician rating sites is still in its beginnings, and a too-specific framework (eg, specific eHealth ethics frameworks [19,20]) runs the risk of excluding relevant ethical issues and arguments, and (2) because the structure provided here aims to inform health care professionals and health policy decision makers that are not always trained in ethics, it is helpful to frame the analysis with well-known ethical principles.

The three substantial ethical principles are (1) patient (and physician) welfare, (2) patient autonomy, and (3) social justice. The analysis of these substantial ethical principles is complemented by a brief description of procedural criteria that aims to improve the fairness of the health policy decision making concerning physician rating sites.

Physician Rating Sites and Patient Welfare

Assuming that there are outcome-related quality differences between physicians and that physician rating sites can detect these differences to a certain degree, then physician rating sites could aim to improve health outcomes (patient welfare). One way to realize this goal is to make doctors aware of aspects of their work in need of improvement as identified by physician rating sites so that adjustments can be implemented in practice [14]. Second, patients who obtain information from physician rating sites are probably more likely to choose better-quality physicians and reap health benefits as a result. However, direct evidence that supports or negates these assumptions is lacking.

Can evidence from other fields be reasonably transferred to physician rating sites? The evidence related to the effects of public reporting of physician performance, based on either specific quality indicators or public report, is mixed [21]. Few researchers have examined the effects of public reporting on the actual delivery of health care [22,23] and even fewer have investigated how report cards might improve care [24,25]. Paradoxically, there is some evidence that public report cards can actually reduce quality by leading physicians to select patients based on risk profiles that best suit the specific quality indicator, which is critical for the interpretation of physician rating sites [26,27]. It is questionable, however, whether this scenario can be transferred to physician rating sites. Nevertheless, opponents of physician rating sites suggest that

XSL•FO
RenderX

patients could be led to award more positive ratings based on nonmedical factors, such as pleasant waiting room music, or even persuasion by the physician.

Although physician rating sites cannot measure positive health outcomes directly, the presence of physician rating sites and the resulting transparency of medical performance could result in a (measurable) improvement in public trust in the system [21,28]. However, evidence either in support of or against this assumption is also lacking.

Physician Rating Sites and Patient Autonomy

Besides the direct improvement in health outcomes, another intervention goal of physician rating sites that is closely linked to the ethical principle *respect of patient autonomy* can be grouped under the heading *improvement of health literacy* [29,30]. According to the World Health Organization's definition, health literacy is "the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand, and use information in ways which promote and maintain good health" [31]. Improving health literacy empowers and thereby respects patient autonomy in making various health-related choices. Physician rating sites could potentially influence health literacy on three different levels [29] (see Table 1 [29,32]).

Table 1. The potential impact of physician rating sites on the three levels of health literacy

Level	Potential impact
Functional	People are able to process simple information that can help them find a specialist physician who understands medical procedures. Obviously, physician rating sites could promote this functional level of health literacy by providing a wide range of information; conversely, they could counteract it by disseminating false or at least biased information (for example, by putting only those physicians at the top if the list who bought a premium account offered by the specific physician rating sites).
Interactive	Through the development of enhanced cognitive and social skills and structures, this level of health literacy allows patients to play a more active role in interactions with their health environments. Physician rating sites could improve this interactive level—for example, if physician rating sites were to serve as a navigational tool with which patients are better able to steer through the health care system and enhance their communication and exchange of knowledge about specific physicians (or hospitals) from peer to peer (for example, by offering open text options at physician rating sites that allow users to describe in a more narrative style how they experienced the performance of a certain physician).
Critical	People have the ability to question so-called standards and to critically evaluate health-related information [29]. It would be practically relevant if, in general, the exchange of information between patients (peer-to-peer) regarding specific physicians and health care facilities would lead to the development of an increasingly higher and more discriminating level of quality assessment of health care through patients themselves. For example, physician ratings could cover aspects of health care quality that other patients have not sought before, thus providing the possibility to expand patients' horizons in terms of quality assessment. Furthermore, users of physician rating sites could post ratings of which physician reviews were more or less helpful to them or may have even been false and misleading. See, for example, the concept of labeling—that is, to provide information about information (meta-information), which can be either descriptive or evaluative [32].

Physician Rating Sites and Physician Welfare

Alongside the consideration of potential benefits and harms of physician rating sites for patients, the process of ethical decision making should also address the possible side effects for physicians. In particular, it should take into account the possible emotional and psychological burdens for physicians, as well as potential financial damages to those physicians who are not rated as well. In addition to the concern of defamation of individual physicians, another broader concern arises that physicians are discussed and furthermore rated not only based on their professional skills but also as a person. Refer to the assessment from the President of the German Medical Association, Frank U Montgomery, that "The only people who speak up on the Internet are those with an extreme opinion" (translation by the author) [15]. Buckman (see above) pointed his arguments in the same direction. Whether the fears of physician representatives are justified is more than questionable. Recent evaluations of rating tendencies from the United States and Germany demonstrate that the majority of reviews in physician rating sites appear to be positive [1,33].

Nevertheless, the potential harm to physicians should be taken seriously. For example, making the ratings first available when they have reached a certain baseline number (eg, 5–10) reduces the impact of extreme opinions, and peer review allows for the differentiation and elimination of defamations. However, when the baseline number or the peer-review requirements are set too high, then potential benefits for patients could be hindered. An ethical and regulative challenge is depicting the differentiation between disproportionate defamation on the one hand and relevant critique on the other hand. The criteria that physician rating sites use for these differentiation tasks (including examples of ratings classified as *defamation* or *relevant critique*) should be made transparent to the users. Furthermore, eliminating overly negative ratings without eliminating overly positive ratings will create a general bias toward more positive

XSL•FO
RenderX

ratings and will decrease the differentiation between physicians. See also the section below on legitimacy of decision making in the implementation and regulation of physician rating sites.

Physician Rating Sites and Social Justice

If we accept the assumption, as discussed above, that physician rating sites could have a positive effect on patient welfare as well as on the advancement of health literacy, then they could also have an impact on equal opportunity among the different socioeconomic groups within one society [30,34]. For reasons of equity, one should investigate whether all socioeconomic groups that could benefit from physician rating sites have unrestricted access to the Internet. The Internet as a source of information regarding the quality of physicians is relatively accessible in comparison with alternative forms of media (print media and personal contact). The relativity arises as a result of the contrast between possible effective alternatives. Arguably, one of the most effective available options to find a good physician is to ask friends or relatives in the medical profession to recommend a colleague. Thus, it is indisputable that physicians as a social group have structural advantages in the search for a new physician due to insider information received from colleagues.

Even though the Internet is widely accessible, one must consider remaining financial and nonfinancial access barriers, such as income, culture, gender, and age. Effective use of physician rating sites remains primarily dependent on the cognitive and intellectual capabilities of the users. This phenomenon could negatively affect the already-existing health inequalities between more- and less-educated groups (inequity).

A further issue to be considered are effects that have been observed in the context of public reporting of quality indicators [21]. If quality indicators such as satisfaction with care are correlated with race and socioeconomic status [35,36], then physicians may shy away from treating some groups of patients out of fear of being penalized by relatively poor ratings in physician rating sites.

Legitimacy of Decision Making in the Implementation and Regulation of Physician Rating Sites

Decisions regarding the implementation or regulation of physician rating sites through public institutions (eg, NHS Choices in the United Kingdom or statutory health insurers in Germany) are associated with inevitably complex issues. Such issues cannot be solved by reference to an ultimate ethical principle [16]. Whenever health care institutions are confronted with the challenges of ethical assessments, they increase the legitimacy of their decisions when certain minimal requirements for a fair decision-making process are met [34] (see Table 2 [37-39]).

Table 2. Basic conditions for a fair decision-making process concerning the implementation and regulation of physician rating sites

Condition	Implication
Transparency	Empirical information and normative arguments that were relevant for the decision-making process on more- or less-restrictive regulation of physician rating sites should be made available to the public.
Justification	Decisions should be based on a relevant rationale. Relevant reasons are especially those that explicitly and comprehensibly ascribe to the above-described ethical criteria: patient and physician welfare, autonomy, and justice.
Participation	Subjective evaluations that are part of the decision-making process are inevitable due to the complexity of the question. The legitimacy of such subjective evaluations increases when the affected populations (here patients, physicians, and insurance agents) have been given the opportunity to participate and to provide relevant empirical information and normative arguments [37,38].
Minimizing conflicts of interest	Decisions on the implementation or regulation of physician rating sites should be regulated in order to avoid as many conflicts of interest as possible [39]. Conflicts of interest exist, for example, if the decision maker him- or herself benefits from any financial advantages on decisions made for or against any particular forms of regulation of physician rating sites.
Openness for revision	Every decision should be open for revision provided that better normative arguments or better evidence on the effects of physician rating sites is available.

Discussion

The previous sections specified fundamental ethical principles relevant to the discussion of the basic concept of physician rating sites (allowing patients to rate, comment on, and discuss physicians' performance—online and visible to everyone). The specified ethical principles should be recognized when the various stakeholders in the field of physician rating sites (physicians, patients, politicians, insurance companies, owners of private physician rating sites, and others) develop their viewpoints toward the basic concept of physician rating sites. These principles should also play a crucial role when decisions on the implementation and (more- or less-restrictive) regulation of physician rating sites are made. Even when thorough empirical evaluations of potential unknown effects of physician rating sites are strongly required, drawing on plausibility and normative arguments is unavoidable for justifying (regulatory) decisions regarding physician rating sites. The aforementioned basic conditions for a fair decision-making process are particularly relevant under such conditions of normative complexity and insufficient evidence (uncertainty).

In the opinion of the author, two issues present a special challenge and should play an important role when deciding about more- or less-restrictive physician rating sites regulations.

XSL•FO
RenderX

First, the potential psychological and financial harms to physicians need to be contained without limiting the potential health and health literacy benefits for patients. In many countries the medical profession enjoys privileges such as strong advocacy groups and special social facilities. Thus, the denial of transparency on patient experiences and satisfaction (with physician performance) requires a strong rationale. Second, in light of the unequal distribution of health opportunities, particularly due to discrepancies in health literacy, possible countermeasures (such as physician rating sites) are highly relevant. Measures to improve public physician rating sites (such as NHS Choices and the AOK website) should be specifically tailored to the needs of vulnerable subgroups of the population. Preferably, aspects such as accessibility and the clarity of information should be evaluated and further improved.

If more general health policy discussions on the acceptance or rejection of the basic ideas of physician rating sites have been settled, further analyses need to focus on the ethical aspects of adequate implementation and management of such websites. Issues such as data privacy and validity, as well as ethical guidelines such as the e-Health Code of Ethics, will then play an important role [19,20,32].

Conflicts of Interest

None declared

References

1. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946. [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
2. Reimann S, Strech D. The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of English- and German-language sites. BMC Health Serv Res 2010;10:332 [FREE Full text] [doi: 10.1186/1472-6963-10-332] [Medline: 21138579]
3. Lagu T, Lindenauer PK. Putting the public back in public reporting of health care quality. JAMA 2010 Oct 20;304(15):1711-1712. [doi: 10.1001/jama.2010.1499] [Medline: 20959582]
4. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
5. Schaefer C, Schwarz S. [Doctor rating sites: which of them find the best doctors in Germany?]. Z Evid Fortbild Qual Gesundhwes 2010;104(7):572-577. [doi: 10.1016/j.zefq.2010.09.002] [Medline: 21095610]
6. Lopez A, Detz A, Ratanawongsa N, Schillinger D, Sarkar U. What do your patients say about you on the internet? A thematic analysis of online reviews of primary care physicians. 2010 Presented at: 33rd Annual Meeting of the Society of General Internal Medicine Meeting; April 28-May 1, 2010; Minneapolis, MN, USA.
7. Black EW, Thompson LA, Saliba H, Dawson K, Black NM. An analysis of healthcare providers' online ratings. Inform Prim Care 2009;17(4):249-253. [Medline: 20359403]
8. Bacon N. Will doctor rating sites improve standards of care? Yes. BMJ 2009;338:b1030. [Medline: 19293223]
9. McCartney M. Will doctor rating sites improve the quality of care? No. BMJ 2009;338:b1033. [Medline: 19293224]
10. Strech D. [Physician rating sites from an ethical viewpoint: a roadmap]. Z Evid Fortbild Qual Gesundhwes 2010;104(8-9):674-681. [doi: 10.1016/j.zefq.2010.09.019] [Medline: 21129706]
11. Carvel J. The Guardian. 2008 Dec 30. Patients to Rate and Review Their GPs on NHS Website URL: http://www.guardian.co.uk/society/2008/dec/30/doctors-rating-website-nhs [accessed 2011-07-29] [WebCite Cache ID 60Y6fhTf7]
12. Jha AK, Orav EJ, Zheng J, Epstein AM. Patients' perception of hospital care in the United States. N Engl J Med 2008 Oct 30;359(18):1921-1931 [FREE Full text] [doi: 10.1056/NEJMsa0804116] [Medline: 18971493]
13. Winefield HR, Murrell TG, Clifford J. Process and outcomes in general practice consultations: problems in defining high quality care. Soc Sci Med 1995 Oct;41(7):969-975. [Medline: 8545671]
14. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. Med Teach 2006 Mar;28(2):117-128. [doi: 10.1080/01421590600622665] [Medline: 16707292]
15. BÄK Newsletter. Ärztlicher Kreisverband Ebersberg Körperschaft des öffentlichen Rechts. 2009 Jun 25. Arztbewertungsportale: massive Kritik am AOK-Arzt-Navigator URL: http://www.aekv-ebersberg.de/aktuelles/105-baek-newsletter-2009-004-vom-25062009 [accessed 2011-07-29] [WebCite Cache ID 60Y6mdCCt]
16. Dyer KA. Ethical challenges of medicine and health on the Internet: a review. J Med Internet Res 2001;3(2):E23 [FREE Full text] [doi: 10.2196/jmir.3.2.e23] [Medline: 11720965]
17. Medical Professionalism Project. Medical professionalism in the new millennium: a physicians' charter. Lancet 2002 Feb 9;359(9305):520-522. [doi: 10.1016/S0140-6736(02)07684-5] [Medline: 11853819]
18. Beauchamp TL, Childress JF. Principles of Biomedical Ethics. 6th edition. New York, NY: Oxford University Press; 2009.
19. Eysenbach G. Towards ethical guidelines for e-health: JMIR theme issue on eHealth ethics. J Med Internet Res 2000;2(1):E7 [FREE Full text] [doi: 10.2196/jmir.2.1.e7] [Medline: 11720926]

XSL•FO
RenderX

20. e-Health Ethics Initiative. e-Health Code of Ethics (May 24). J Med Internet Res 2000;2(2):E9 [FREE Full text] [doi: 10.2196/jmir.2.2.e9] [Medline: 11720928]
21. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. JAMA 2005 Mar 9;293(10):1239-1244. [doi: 10.1001/jama.293.10.1239] [Medline: 15755946]
22. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. JAMA 2000 Apr 12;283(14):1866-1874 [FREE Full text] [Medline: 10770149]
23. Leatherman ST, Hibbard JH, McGlynn EA. A research agenda to advance quality measurement and improvement. Med Care 2003 Jan;41(1 Suppl):I80-I86. [Medline: 12544819]
24. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. Med Care 2003 Jan;41(1 Suppl):I30-I38. [Medline: 12544814]
25. Schauffler HH, Mordavsky JK. Consumer reports in health care: do they make a difference? Annu Rev Public Health 2001;22:69-89. [doi: 10.1146/annurev.publhealth.22.1.69] [Medline: 11274512]
26. Omoigui NA, Miller DP, Brown KJ, Annan K, Cosgrove D, Lytle B, et al. Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes. Circulation 1996 Jan 1;93(1):27-33 [FREE Full text] [Medline: 8616936]
27. Werner RM, Asch DA, Polsky D. Racial profiling: the unintended consequences of coronary artery bypass graft report cards. Circulation 2005 Mar 15;111(10):1257-1263 [FREE Full text] [doi: 10.1161/01.CIR.0000157729.59754.09] [Medline: 15769766]
28. Blendon RJ, DesRoches CM, Brodie M, Benson JM, Rosen AB, Schneider E, et al. Views of practicing physicians and the public on medical errors. N Engl J Med 2002 Dec 12;347(24):1933-1940 [FREE Full text] [doi: 10.1056/NEJMsa022151] [Medline: 12477944]
29. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
30. Kickbusch I. Health literacy: engaging in a political debate. Int J Public Health 2009;54(3):131-132. [doi: 10.1007/s00038-009-7073-1] [Medline: 19370309]
31. World Health Organization. Health Promotion Glossary. Geneva: WHO; 1998. URL: http://www.who.int/hpr/NPH/docs/hp_glossary_en.pdf [accessed 2011-11-18] [WebCite Cache ID 63IVjHxT5]
32. Eysenbach G, Diepgen TL. Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. BMJ 1998 Nov 28;317(7171):1496-1500 [FREE Full text] [Medline: 9831581]
33. Strech D, Reimann S. [German language physician rating sites: the status quo of evaluation criteria, evaluation tendencies and utilization]. Gesundheitswesen 2011;73:1-8. [doi: 10.1055/s-0031-1297254]
34. Daniels N. Just Health: Meeting Health Needs Fairly. New York, NY: Cambridge University Press; 2008.
35. Harpole LH, Orav EJ, Hickey M, Posther KE, Brennan TA. Patient satisfaction in the ambulatory setting. Influence of data collection methods and sociodemographic factors. J Gen Intern Med 1996 Jul;11(7):431-434. [Medline: 8842937]
36. Fiscella K, Franks P. Influence of patient education on profiles of physician practices. Ann Intern Med 1999 Nov 16;131(10):745-751. [Medline: 10577297]
37. Nilsen ES, Myrhaug HT, Johansen M, Oliver S, Oxman AD. Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. Cochrane Database Syst Rev 2006(3):CD004563. [doi: 10.1002/14651858.CD004563.pub2] [Medline: 16856050]
38. Thompson AG. The meaning of patient involvement and participation in health care consultations: a taxonomy. Soc Sci Med 2007 Mar;64(6):1297-1310. [doi: 10.1016/j.socscimed.2006.11.002] [Medline: 17174016]
39. Lo B, Field MJ, Committee on Conflict of Interest in Medical Research, Education, and Practice, Institute of Medicine. Conflicts of Interest in Medical Research, Education, and Practice. Conflicts of Interest in Medical Research, Education, and Practice. Washington, DC: National Academies Press, Institute of Medicine; 2009.

Abbreviations

NHS: National Health Service

Edited by G Eysenbach; submitted 29.07.11; peer-reviewed by T Lagu, E Black, P Wicks, A Datta; comments to author 31.08.11; revised version received 21.10.11; accepted 25.10.11; published 06.12.11.

Please cite as:
Strech D
Ethical Principles for Physician Rating Sites
J Med Internet Res 2011;13(4):e113
URL: http://www.jmir.org/2011/4/e113/
doi:10.2196/jmir.1899
PMID:22146737

XSL•FO
RenderX

©Daniel Strech. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Dashboard Design Using Multiple Data Streams for Disease Surveillance With Influenza Surveillance as an Example

Calvin KY Cheng[1], MMedSc; Dennis KM Ip[1], MBBS, MPhil; Benjamin J Cowling[1], PhD; Lai Ming Ho[1], PhD; Gabriel M Leung[1], MD, MPH; Eric HY Lau[1], PhD

School of Public Health, The University of Hong Kong, Hong Kong, China

Corresponding Author:
Eric HY Lau, PhD
School of Public Health
The University of Hong Kong
5/F William MW Mong Block, Faculty of Medicine Building, 21 Sassoon Road,
Hong Kong
China
Phone: 852 3906 2019
Fax: 852 3520 1944
Email: ehylau@hku.hk

Abstract

Background: Great strides have been made exploring and exploiting new and different sources of disease surveillance data and developing robust statistical methods for analyzing the collected data. However, there has been less research in the area of dissemination. Proper dissemination of surveillance data can facilitate the end user's taking of appropriate actions, thus maximizing the utility of effort taken from upstream of the surveillance-to-action loop.

Objective: The aims of the study were to develop a generic framework for a digital dashboard incorporating features of efficient dashboard design and to demonstrate this framework by specific application to influenza surveillance in Hong Kong.

Methods: Based on the merits of the national websites and principles of efficient dashboard design, we designed an automated influenza surveillance digital dashboard as a demonstration of efficient dissemination of surveillance data. We developed the system to synthesize and display multiple sources of influenza surveillance data streams in the dashboard. Different algorithms can be implemented in the dashboard for incorporating all surveillance data streams to describe the overall influenza activity.

Results: We designed and implemented an influenza surveillance dashboard that utilized self-explanatory figures to display multiple surveillance data streams in panels. Indicators for individual data streams as well as for overall influenza activity were summarized in the main page, which can be read at a glance. Data retrieval function was also incorporated to allow data sharing in standard format.

Conclusions: The influenza surveillance dashboard serves as a template to illustrate the efficient synthesization and dissemination of multiple-source surveillance data, which may also be applied to other diseases. Surveillance data from multiple sources can be disseminated efficiently using a dashboard design that facilitates the translation of surveillance information to public health actions.

(J Med Internet Res 2011;13(4):e85) doi:10.2196/jmir.1658

KEYWORDS

Dashboard; dissemination; surveillance; influenza

Introduction

Respiratory viruses cause significant global mortality and morbidity each year. Influenza virus is of particular public health concern due to its association with severe infections and deaths [1,2]. Disease surveillance provides useful information that helps monitoring trends and disease burden, planning, implementing, and evaluating appropriate prevention and control interventions as well as allocating resources [3]. At present, the World Health Organization (WHO) runs the largest human influenza virus surveillance network [4], which includes 136 institutions from 106 countries mainly focusing on the genetic and antigenic characteristics of prevailing strains. While the original purpose of WHO's global influenza surveillance

XSL•FO
RenderX

network was to recommend the content of the influenza vaccine for the subsequent influenza season [5], local and national prospective influenza and influenza-like illness surveillance systems also provide important information to policy makers and public health practitioners for situational awareness during periods of influenza activity.

Great strides have been made exploring and exploiting new and different sources of respiratory disease data, especially after the severe acute respiratory syndrome (SARS) epidemic in 2003 [6,7] and the 2009 influenza A (H1N1) pandemic. However, as different surveillance methods target only specific target groups of population possibly at different stages of disease progression, thus providing related but different information [8], using a single surveillance data stream to reflect overall disease activity may not be appropriate. It remains challenging to incorporate multiple surveillance data and fully utilize the available information from the data to summarize the overall situation.

On the other hand, there has been rather less research in the area of data dissemination, while the feedback of contextualized data to stakeholders is and must be a key aspect of any surveillance system if confidence with and enthusiasm for the system is to be maintained [9,10]. It is also important that the dissemination is not neglected. No matter how accurate and timely the upstream surveillance data collection and analysis of the surveillance system are, the return for the time and effort invested would be discounted if the means of health communication for data dissemination were suboptimal. Poor communications could lead to delayed information transfer, loss of information, or even misinterpretation of the surveillance results. Effort should also be taken for data dissemination to complete the surveillance-to-action loop [3].

Digital dashboards (also known as executive dashboards) describe computerized interactive tools typically used by managers to visually ascertain the status (or "health") of their business via key performance indicators. These tools emerged from the concepts of decision support systems in the 1970s, and, with the rapid growth in information technology through the 1990s, these have developed into standard tools in executives offices [11,12]. Their use has expanded to the field of medical science, for example, displaying bed occupancy, the availability of clinical staffs for hospital management [13,14], or monitoring the stock of different pharmaceutical items in the dispensary [15]. In a more general sense, digital dashboards simply describe systems for the visual presentation of key indicators in appropriate context, allowing rapid interpretation by an end user. A dynamic dashboard can be made interactive and user-friendly by allowing extraction of information in different perspectives and contexts. By investing in the design of data dissemination, the quality of presentation and communication of data by local health authorities as well as the general public could be much improved, thus maximizing the usefulness of available data of the surveillance systems.

In this research paper, we developed a generic framework for a digital dashboard and illustrate this framework by specific application to influenza surveillance. Specific features were designed and incorporated according to the requirements identified to facilitate efficient dissemination of surveillance data. The dashboard can reveal both the influenza activity in different population sectors and the overall situation at a glance, maximizing the use of available data for policy makers as well as the general public for appropriate actions.

Methods

Rationale and Implementation of the Dashboard Design

We designed an online interactive dashboard that fit seamlessly with human visual perception. Three authors of the current study (CKYC, BJC, EHYL), with backgrounds in computer science, epidemiology, and statistics, identified weaknesses of current surveillance data dissemination methods according to a previous review study [8]. From this review and from experience of and feedback from the 2009 influenza pandemic, we identified the requirements for the new system. In response to these requirements, the design of the dashboard features incorporated merits of the national surveillance websites identified in the earlier the review, applied principles of efficient data presentation and dashboard design [16,17], and adopted recommendations from professional information technology consultants. Specific examples of each component were mainly extracted from the influenza surveillance dashboard, which was constructed as a demonstration of these design principles.

In the main webpage and drill-down pages, we implemented the following features for efficient data presentation: (1) provision of information that viewers need quickly and clearly, (2) organization of information to support meaning and usability, (3) minimization of distractions, clichés, and unnecessary embellishments that could create confusion, (4) creation of an aesthetically pleasing viewing experience, and (5) consistency of design for easy data comparison.

The dashboard displays and synthesizes surveillance data at one glance to provide an overview of disease activity from different data streams. Scrolling of pages was avoided so that users can easily focus on the information from different sources rather than searching for information across different pages or sections. All surveillance raw data were stored in standard data file format to achieve data manipulation simplicity and reduce data processing time during the update and server queries of the dashboard. Several pages allow interactive graphics and information display according to end users' instruction, and we also allow raw data export in common standard data file formats.

Data Used for Influenza Surveillance

We used five different types of influenza surveillance data for demonstration of the dashboard, namely, the weekly consultation rate of influenza-like illness reported by general out-patient clinics in Hong Kong, weekly consultation rate of influenza-like illness reported by general practitioners (GPs), weekly influenza virus isolation rate, weekly overall school absenteeism rate, and weekly hospital admission rate of children aged 4 and under with principal diagnosis of influenza. The school absenteeism data was provided by BroadLearning Education (Asia) Ltd, an online electronic school administration system service provider. The other four data streams were extracted from the official

XSL•FO
RenderX

website of Centre for Health Protection, Department of Health of the Hong Kong Special Administrative Government [18].

System Development

The system was designed to allow continual development such as incorporating more surveillance data streams or applying more sophisticated analytical algorithms for aberration detection by modifying specific related components of the internal R programs. Little additional programming was needed for database storage and graphic display with changes in surveillance data sources, for example, addition of data from a new surveillance system.

Results

The influenza surveillance dashboard is hosted on the server of the School of Public Health, the University of Hong Kong (http://sph.hku.hk/dashboard). Here we present the design patterns [19,20] describing the main features of the surveillance dashboard.

System Architecture

Requirements

The requirements are that the system be simple, stable, and require few resources. An additional requirement is for efficient data transfer and analysis to facilitate timely surveillance of disease patterns.

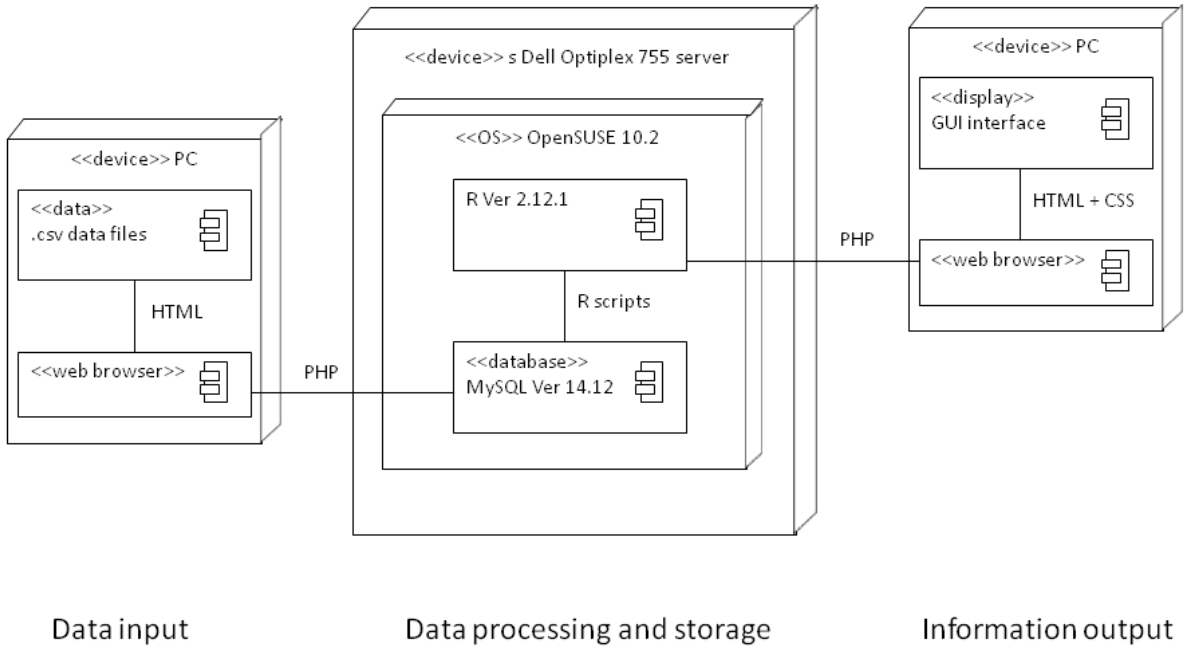
Design

A simple 3-tiered structure with database, logical, and presentation layers was adopted. The system was built on open source programming languages. Command line based analyzing software with streamline programming was incorporated to avoid unnecessary computational procedures.

Example

For the presentation layer, we used hypertext markup language (HTML) together with cascading style sheets (CSS) and hypertext preprocessor (PHP) Web programming language. For the back-end database, we used the MySQL (structured query language) (Oracle Corporation, Redwood Shores, CA) server. Data in comma separated value (CSV) format were imported to the MySQL server via HTML. The logical layer for data analysis and generating graphical results is based on R version 2.12.1 (R Development Core Team, Vienna, Austria). Figure 1 illustrates the diagrammatic presentation of the system designed in unified modelling language (UML) 2. In this system, each component was connected by specific programming languages via two-way communication.

Figure 1. Deployment diagram of the surveillance dashboard design.



XSL•FO
RenderX

Data Input, Storage, and Extraction

Requirements

A requirement was automated data update to minimize operating resources and human errors. Also required were compatible formats to facilitate data sharing and manipulation in different operating systems and programs.

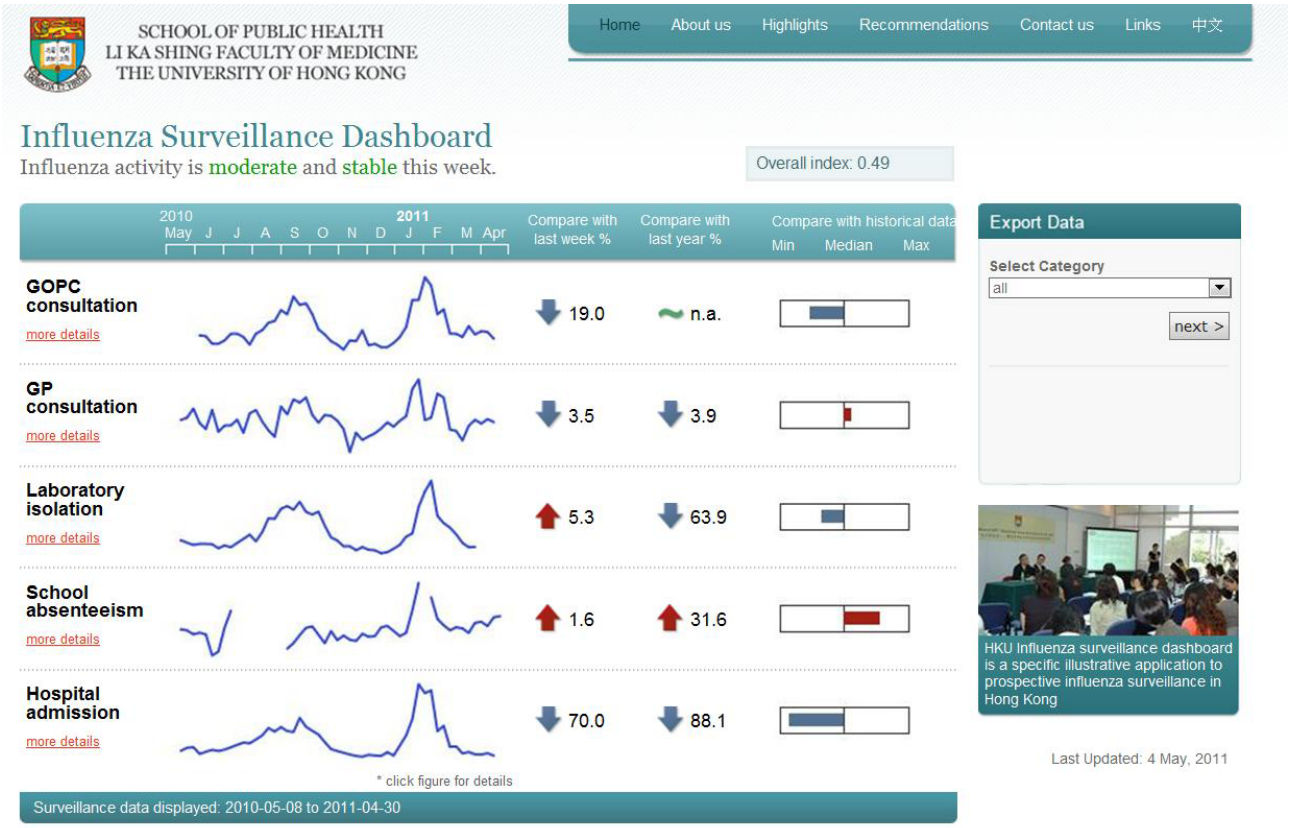
Design

Format of all raw data were standardized for input and extraction. Surveillance data were stored in a central database connecting to all webpages for easy update. Raw data can be extracted from the database using standard query language (SQL) according to the users' specification in a user-friendly menu embedded in the webpage.

Example

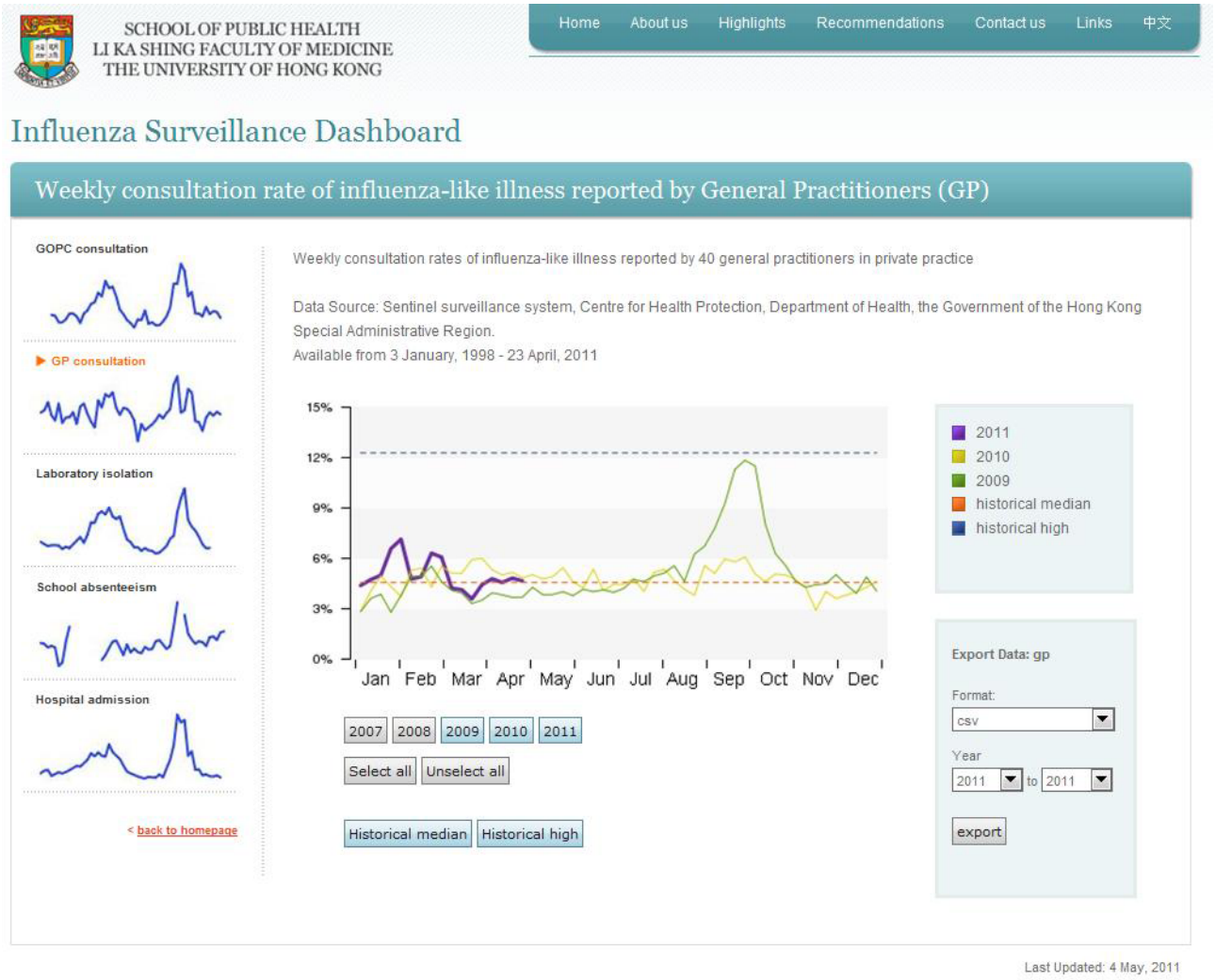
The database can be imported by the one-step data administration system named the "importer," a tailor-made online data file import system that has been linked to the Internet browser, the back-end server database, and internal PHP and R programs. Once the updated raw data file is ready, the administrative personnel simply upload this file at the importer webpage. The importer will automatically detect the file format and update the central backend MySQL server data in the corresponding data tables. The sparkline and time series image graphs (jpeg format) at the main webpage and drill-down pages will also be updated by the R programs. For data extraction, end users may select specific data stream(s) as well as the time period to extract raw data for their own specific purposes at the main webpage and drill-down pages in comma separated value (CSV) or extensible markup language (XML) format (the panel on the right in Figures 2 and 3).

Figure 2. Screenshot of the main page of the influenza surveillance dashboard.



XSL•FO
RenderX

Figure 3. Screenshot of a drill-down page showing the individual data stream (GP consultations) in the influenza surveillance dashboard.



Presentation Layer

Requirements

Requirements are display of all surveillance data at one glance in the dashboard to provide a clear overview of disease activity and an interactive display according to users' specific needs. Also required are graphical alerts to highlight elevated disease activity to achieve effective communications of health information.

Design

Webpages were designed to be content-rich and self-explanatory to present surveillance data streams in clear and simple figures. Distractions were avoided so that surveillance data would stand out on the page. Bright colours or thicker lines were used to indicate high levels or more recent data. Surveillance data streams with increasing trend or at a high level were highlighted by indicators. As surveillance data streams reflect disease activity in different target populations, this allows users to identify elevated disease activity in a certain population. For dynamic graphical display, specific parameters selected by the user in the Web graphical user interface (GUI) will be transferred to the internal programs for generating the most updated figures. Previous and outdated figures will be cleaned up to avoid wastage of server memory.

Example

Main Page

A screenshot of the influenza surveillance dashboard main page is shown in Figure 2. In this figure, the five influenza surveillance data streams in the last 12 months are shown in panels to allow quick comparison of disease trends. On the left, brief titles for the data streams are given where more details will be shown when they are pointed to. Based on the back-end algorithm, the level (low, moderate, or high) and trend (decreasing, stable, or increasing) of the overall influenza activity are shown on the top left. The overall influenza activity index generated by the dynamic linear model ranging from 0 to 1 is shown in the top right. Different formats of raw data can be extracted using the drop-down menu on the right.

Other measurements or indicators such as comparisons with past data can be added to provide supplementary information of the individual data streams. Raw data of selected data streams and years can also be exported using the drop-down menu on the right.

Individual Surveillance Data Stream Drill-Down Pages

The drill-down page for the corresponding individual data stream will be displayed by clicking the figure for the individual data stream (Figure 3). The drill-down page provides more

XSL•FO
RenderX

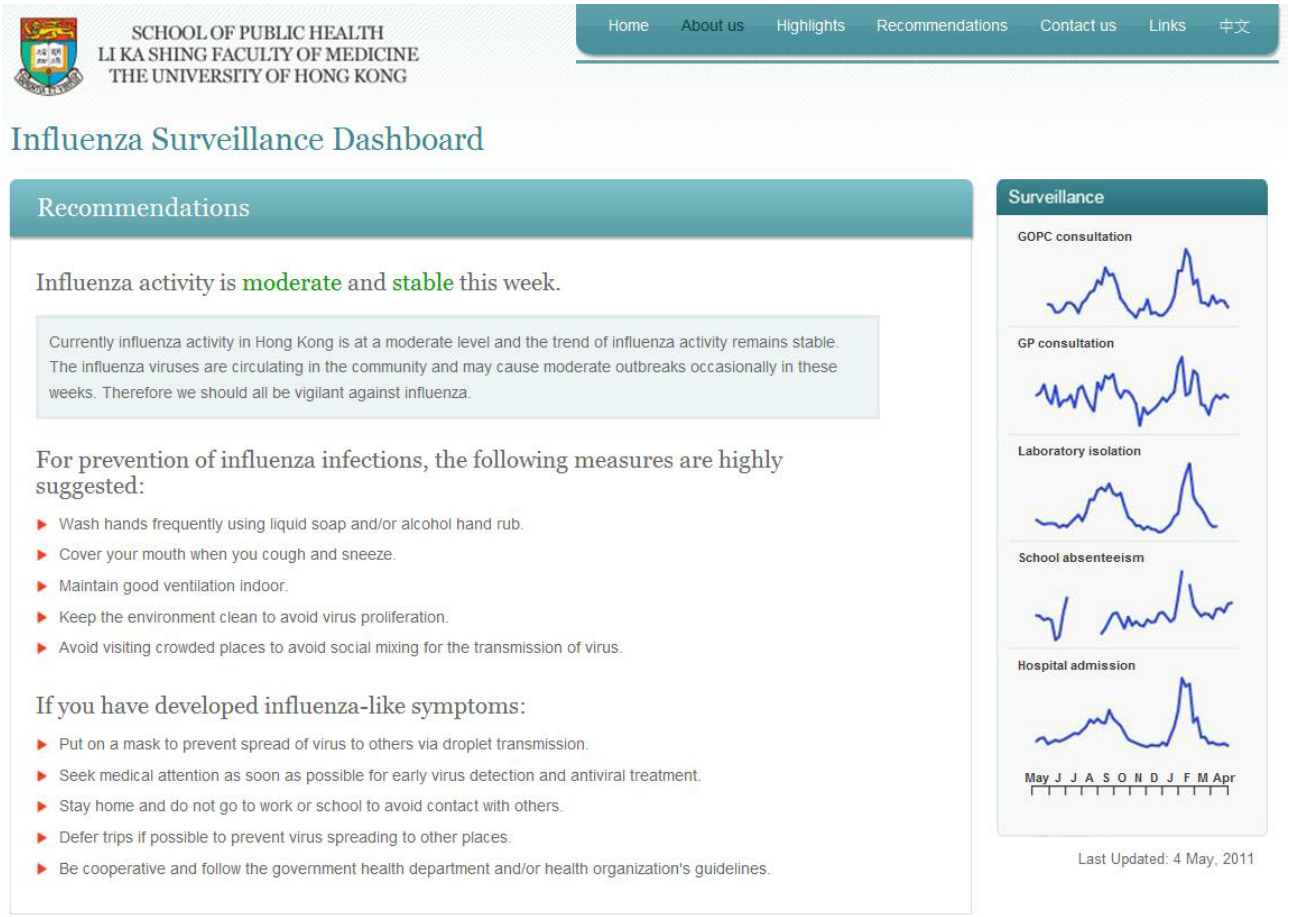
detailed information, such as data from past years, the historical median, or the high level. Users can toggle specific information to be displayed by clicking the corresponding buttons. A thicker line is used to show data in the current year. A navigator bar on the left allows users to select different surveillance data streams. Further stratification of the data can be incorporated in the drill-down page whenever the data are available. We provided stratification of school absenteeism data by district for illustration.

Recommendations Drill-Down Page

As shown in Figure 4, the "Recommendations" drill-down page provides the general public different recommendations and motivates actions for prevention of influenza infection and transmission corresponding to the level and trend of overall influenza activity.

Other drill-down pages provided supplementary information on the sources of the data streams and other related information. The Chinese version of the influenza surveillance dashboard can be assessed by clicking the "Chinese" option at the top right of each page.

Figure 4. Screenshot of the Recommendations drill-down page in the influenza surveillance dashboard.



Discussion

We developed an influenza surveillance dashboard that serves as a template to demonstrate efficient synthesization and dissemination of multiple-source surveillance data, which may also be applied to other diseases. Surveillance data from multiple sources can be disseminated efficiently using a dashboard design, which facilitates the translation of surveillance information to public health actions.

Public health surveillance is defined by Thacker as "the ongoing, systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination to those who need to know" [21]. Previously, much effort and many resources were required to improve the quality of surveillance data and the accuracy of diagnostic tests. Until the late 1990s, with the rapid development of information technology and the Internet, the timeliness of surveillance data transfer improved. Nevertheless, the methods and targets of data dissemination have been somewhat neglected, probably because communication and graphic design are apparently unrelated to the field of medical science. Indeed, timely and effective health communication is extremely valuable to transfer information to the relevant people who can take appropriate action. In particular, timeliness is always a critical factor for infectious disease containment and control. Comparatively speaking, fewer resources are needed to develop effective data dissemination methods than are required for data collection, quality control, diagnosis, and analysis. Researchers should also keep an eye on data dissemination to maximize the usefulness of the upstream surveillance effort.

XSL•FO
RenderX

A review of national surveillance websites [8] provided evidence that less attention has been paid to designing user-friendly and efficient dissemination of surveillance data. In particular, many websites have presented or highlighted data in report form, which made information more difficult to be identified or synthesized. Long reports with scattered information over pages also made it difficult to grasp the overall situation of disease activity. Also, functions for data retrieval were not commonly available for most websites. Such functions are greatly needed for efficient and timely data sharing.

The dashboard style can automatically summarize data and present figures of different types in the drill-down pages. Other detailed or supplementary information, such as past records of individual surveillance data streams or recommendations can be available in other drill-down pages. We also incorporated a data export function in standard format (eg, CSV and XML) to facilitate data sharing, which was shown to be important for epidemiological enquiry in the recent human influenza A (H1N1) pandemic.

Moreover, a primary motivation for publicizing surveillance data online should be to allow timely risk communication thereby facilitating disease prevention [22]. As public health awareness has been elevated since the SARS outbreak in 2003, it is expected that sources of surveillance data will only increase. In the influenza surveillance dashboard, we demonstrated how recommendations can be made corresponding to multiple surveillance data so that the collected information can be more efficiently transferred into public health actions.

While more surveillance data streams are being developed, there is a need for developing more sophisticated algorithms for aberration detection. Here, some challenges remain for the development of multivariable analytic algorithms as different surveillance data streams represent situations of different targets at different stages of disease progression. Furthermore, evidence is still lacking about the association between these new surveillance data streams and influenza activity. Whether these data provide more signal or noise is still in doubt due to their nonspecific nature. These data streams will still need to be evaluated and filtered to increase the signal-to-noise ratio before adding them as a routine surveillance practice; thus, they can only serve as supplementary information at the current stage. Other challenges for using combined surveillance data from different sources include the handling of missing data (eg, school absenteeism during school holidays), data provided in different resolutions (daily versus weekly resolutions), and unsynchronized data receiving protocols from different sources, which the algorithm should also be capable of handling.

Generally speaking, an ideal surveillance website should have comparable hardware and software and a standard user interface, data format, and coding for easy data sharing. The influenza surveillance dashboard demonstrates how different sources of surveillance data streams can be synthesized and efficiently displayed in a dashboard to provide an overview of disease activity as well as each individual source. The influenza surveillance dashboard serves as a template that can provide for efficient dissemination in a similar manner of information about other diseases with multiple surveillance data sources. With a back-end algorithm, indicators or statistics can be generated from individual or all data streams. Other more complicated indicators can also be generated that are tailored to specific disease type by modifying the back-end algorithm.

Acknowledgments

This research was funded by the Research Fund for the Control of Infectious Diseases of the Health, Welfare and Food Bureau of the Hong Kong SAR Government (grant number 08070662), the Area of Excellence Scheme of the Hong Kong University Grants Committee (grant number AoE/M-12/06), and the US National Institutes of Health Models of Infectious Disease Agent Study program (grant number 1 U54 GM088558). We gratefully acknowledge the Centre for Health Protection, Hong Kong; the Hospital Authority, Hong Kong; and BroadLearning Education (Asia) Ltd for providing influenza surveillance data. We thank Wing-Hong Seto for useful discussion, Xiu-Qing He for research support, and two anonymous reviewers for their comments and suggestions.

Conflicts of Interest

Author DKMI has received research funding from Hoffmann-La Roche Inc. Author BJC has received research funding from MedImmune Inc, a manufacturer of influenza vaccines. The authors report no other conflicts of interest.

References

1. Wong CM, Chan KP, Hedley AJ, Peiris JS. Influenza-associated mortality in Hong Kong. Clin Infect Dis 2004 Dec 1;39(11):1611-1617 [FREE Full text] [doi: 10.1086/425315] [Medline: 15578360]
2. Wong CM, Yang L, Chan KP, Leung GM, Chan KH, Guan Y, et al. Influenza-associated hospitalization in a subtropical city. PLoS Med 2006 Apr;3(4):e121 [FREE Full text] [doi: 10.1371/journal.pmed.0030121] [Medline: 16515368]
3. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN, Guidelines Working Group Centers for Disease Control and Prevention (CDC). Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep 2001 Jul 27;50(RR-13):1-35; quiz CE1. [Medline: 18634202]
4. World Health Organization. WHO Global Influenza Surveillance Network. Geneva, Switzerland: World Health Organization; 2011. URL: http://www.who.int/influenza/surveillance_monitoring/en/ [accessed 2011-09-06] [WebCite Cache ID 62MxdkPh3]

XSL•FO
RenderX

5. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, et al. Influenza vaccine strain selection and recent studies on the global migration of seasonal influenza viruses. Vaccine 2008 Sep 12;26 Suppl 4:D31-D34. [Medline: 19230156]
6. Chan WF, Wong TK. Preparing for pandemic influenza: revisit the basics. J Clin Nurs 2007 Oct;16(10):1858-1864. [doi: 10.1111/j.1365-2702.2007.01831.x] [Medline: 17880474]
7. Foldy SL, Barthell E, Silva J, Biedrzycki P, Howe D, Erme M, et al. SARS Surveillance Project--Internet-enabled multiregion surveillance for rapidly emerging disease. MMWR Morb Mortal Wkly Rep 2004 Sep 24;53 Suppl:215-220 [FREE Full text] [Medline: 15717395]
8. Cheng CK, Lau EH, Ip DK, Yeung AS, Ho LM, Cowling BJ. A profile of the online dissemination of national influenza surveillance data. BMC Public Health 2009;9:339 [FREE Full text] [doi: 10.1186/1471-2458-9-339] [Medline: 19754978]
9. Cameron R, Manske S, Brown KS, Jolin MA, Murnaghan D, Lovato C. Integrating public health policy, practice, evaluation, surveillance, and research: the school health action planning and evaluation system. Am J Public Health 2007 Apr;97(4):648-654 [FREE Full text] [doi: 10.2105/AJPH.2005.079665] [Medline: 17329662]
10. Grant A, Moshyk A, Diab H, Caron P, de Lorenzi F, Bisson G, et al. Integrating feedback from a clinical data warehouse into practice organisation. Int J Med Inform 2006;75(3-4):232-239. [doi: 10.1016/j.ijmedinf.2005.07.037] [Medline: 16153886]
11. Nagy PG, Konewko R, Warnock M, Bernstein W, Seagull J, Xiao Y, et al. Novel, Web-based, information-exploration approach for improving operating room logistics and system processes. Surg Innov 2008 Mar;15(1):7-16. [doi: 10.1177/1553350608316573] [Medline: 18407926]
12. Wadsworth T, Graves B, Glass S, Harrison AM, Donovan C, Proctor A. Using business intelligence to improve performance. Healthc Financ Manage 2009 Oct;63(10):68-72. [Medline: 19810655]
13. Mazzella-Ebstein AM, Saddul R. Web-based nurse executive dashboard. J Nurs Care Qual 2004;19(4):307-315. [Medline: 15535535]
14. Park KW, Smaltz D, McFadden D, Souba W. The operating room dashboard. J Surg Res 2010 Dec;164(2):294-300. [doi: 10.1016/j.jss.2009.09.011] [Medline: 20006345]
15. Bahl V, McCreadie SR, Stevenson JG. Developing dashboards to measure and manage inpatient pharmacy costs. Am J Health Syst Pharm 2007 Sep 1;64(17):1859-1866. [doi: 10.2146/ajhp060596] [Medline: 17724369]
16. Few S. Information Dashboard Design: The Effective Visual Communication of Data. Beijing, China: O'Reilly; 2006.
17. Malik S. Enterprise Dashboards: Design and Best Practices for IT. Hoboken, NJ: Wiley; 2005.
18. Centre for Health Protection, Department of Health.: The Government of Hong Kong Special Administrative Region; 2011. URL: http://www.chp.gov.hk/ [accessed 2011-09-06] [WebCite Cache ID 62MwfXdd4]
19. Lasater CG. Design Patterns (Wordware Applications Library). Piano, TX: Wordware Publishing, Inc; 2007.
20. Timpka T, Eriksson H, Gursky EA, Strömgren M, Holm E, Ekberg J, et al. Requirements and design of the PROSPER protocol for implementation of information infrastructures supporting pandemic response: a Nominal Group study. PLoS One 2011;6(3):e17941 [FREE Full text] [doi: 10.1371/journal.pone.0017941] [Medline: 21464918]
21. Thacker SB. Historical development. In: Lee LM, Teutsch SM, Thacker SB, St Louis ME, editors. Principles and Practice of Public Health Surveillance. 3rd edition. New York, NY: Oxford University Press; 2010.
22. Ackoff RL. From data to wisdom. Journal of Applied Systems Analysis 1989;16:3-9.

Abbreviations

CSS: cascading style sheets
CSV: comma separated value
GP: general practitioner
GUI: graphical user interface
HTML: hypertext markup language
PHP: hypertext preprocessor
SARS: severe acute respiratory syndrome
SQL: standard query language
WHO: World Health Organization
UML: unified modeling language
XML: extensible markup language

XSL•FO
RenderX

Edited by G Eysenbach; submitted 02.10.10; peer-reviewed by D Uglow, T Timpka; comments to author 18.04.11; revised version received 21.06.11; accepted 24.06.11; published 14.10.11.

Please cite as:
Cheng CKY, Ip DKM, Cowling BJ, Ho LM, Leung GM, Lau EHY
Digital Dashboard Design Using Multiple Data Streams for Disease Surveillance With Influenza Surveillance as an Example
J Med Internet Res 2011;13(4):e85
URL: http://www.jmir.org/2011/4/e85/
doi:10.2196/jmir.1658
PMID:22001082

©Calvin KY Cheng, Dennis KM Ip, Benjamin J Cowling, Lai Ming Ho, Gabriel M Leung, Eric HY Lau. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.10.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Attrition in Web-Based Treatment for Problem Drinkers

Marloes G Postel[1,2], PhD; Hein A de Haan[1,2], MD; Elke D ter Huurne[1,2], MSc; Job van der Palen[3,4], PhD; Eni S Becker[5], PhD; Cor AJ de Jong[2,5], MD, PhD

[1]Tactus Addiction Treatment, Enschede, Netherlands

[2]Nijmegen Institute for Scientist-Practitioners in Addiction, Nijmegen, Netherlands

[3]Department of Epidemiology, Medisch Spectrum Twente, Enschede, Netherlands

[4]Department of Research Methodology, Measurement, and Data Analysis, University of Twente, Enschede, Netherlands

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

Corresponding Author:
Marloes G Postel, PhD
Tactus Addiction Treatment
Institutenweg 1
Enschede, 7521 PH
Netherlands
Phone: 31 534824750
Fax: 31 534824760
Email: m.postel@tactus.nl

Abstract

Background: Web-based interventions for problem drinking are effective but characterized by high rates of attrition. There is a need to better understand attrition rates in order to improve the completion rates and the success of Web-based treatment programs.

Objective: The objectives of our study were to (1) examine attrition prevalence and pretreatment predictors of attrition in a sample of open-access users of a Web-based program for problem drinkers, and (2) to further explore attrition data from our randomized controlled trial (RCT) of the Web-based program.

Methods: Attrition data from two groups of Dutch-speaking problem drinkers were collected: (1) open-access participants enrolled in the program in 2009 (n = 885), and (2) RCT participants (n = 156). Participants were classified as noncompleters if they did not complete all 12 treatment sessions (9 assignments and 3 assessments). In both samples we assessed prevalence of attrition and pretreatment predictors of treatment completion. Logistic regression analysis was used to explore predictors of treatment completion. In the RCT sample, we additionally measured reasons for noncompletion and participants' suggestions to enhance treatment adherence. The qualitative data were analyzed using thematic analysis.

Results: The open-access and RCT group differed significantly in the percentage of treatment completers (273/780, 35.0% vs 65/144, 45%, $\chi^2_1 = 5.4$, $P = .02$). Logistic regression analysis revealed a significant contribution of treatment readiness, gender, education level, age, baseline alcohol consumption, and readiness to change to predict treatment completion. The key reasons for noncompletion were personal reasons, dissatisfaction with the intervention, and satisfaction with their own improvement. The main suggestions for boosting strategies involved email notification and more flexibility in the intervention.

Conclusions: The challenge of Web-based alcohol treatment programs no longer seems to be their effectiveness but keeping participants involved until the end of the treatment program. Further research should investigate whether the suggested strategies to improve adherence decrease attrition rates in Web-based interventions. If we can succeed in improving attrition rates, the success of Web-based alcohol interventions will also improve and, as a consequence, their public health impact will increase.

Trial: International Standard Randomized Controlled Trial Number (ISRCTN): 39104853; http://www.controlled-trials.com/ISRCTN39104853 (Archived by WebCite at http://www.webcitation.org/63IKDul1T)

(J Med Internet Res 2011;13(4):e117) doi:10.2196/jmir.1811

KEYWORDS

Web-based treatment; online treatment; problem drinking; alcohol; attrition; dropout

XSL•FO
RenderX

Introduction

Web-based interventions for problem drinkers improve the availability of alcohol treatment services and reach a more diverse segment of the population of problem drinkers [1,2]. Evidence supports the clinical effectiveness of a diversity of Web-based interventions varying from pure self-help to predominantly therapist-administered therapy [3-13], and it seems that the best results are achieved with interventions that use personalized feedback [3]. Despite these promising results, participants in Web-based interventions show great variation in how they use the interventions in terms of frequency and duration of visits, and they often do not complete treatment sessions or assessments [13-16]. However, Web-based intervention studies mainly focus on effectiveness, and less is known about the reasons for noncompletion and the specific components that improve adherence [15]. Although Web-based interventions have the potential of easy data collection, the study of attrition is still rare.

In his law of attrition, Eysenbach distinguished two processes of attrition: dropout attrition and nonusage attrition [15]. Dropout attrition refers to participants being lost to follow-up; they do not return to fill in follow-up questionnaires. Nonusage attrition refers to participants' stopping to use the intervention, but still filling in questionnaires. Authors do not always describe dropout and nonusage attrition separately [4,8-10,17]. And if they do, then much variation is still possible within Eysenbach's conceptualization, because of differences in treatment intervention and study design. Some studies, for example, only require filling out a single questionnaire in a shorter timeframe [4,10], while other studies require a wide range of questionnaires at a number of follow-up points [11,13]. Therefore, comparing attrition rates alone does not make sense. A clear description of the study characteristics, together with nonusage and dropout attrition, is necessary to interpret attrition data properly.

Usage and follow-up completion rates of Web-based alcohol interventions studies published to date range from 16.5% [18] to 92% [11]. In face-to-face addiction treatment, overall around 50% of patients terminate treatment before the intended period is over [19]. The great differences in attrition rates between Web-based interventions can be explained by differences in payment of incentives, follow-up periods, intensity and duration of the Web-based intervention, recruitment procedure, study population, and research environment (trial or open access). It seems to be the combination of factors that is responsible for the attrition rate rather than a single factor [4-6,8-13,18,20-23]. For example, Cunningham et al [11] found that 92% of participants completed baseline, 3-month, and 6-month follow-ups. This excellent follow-up completion rate might partly be explained by the incentive of a $20 check for each follow-up survey, but also by the way participants were recruited. Respondents from an ongoing telephone survey, who confirmed that they had home Internet access and were interested in a computerized program to check their drinking, were invited to participate in the study and therefore motivated respondents were recruited. Doumas and colleagues [10] also found a very good follow-up completion rate (88%) even without paying an incentive. However, their follow-up period was short, at 30 days, and the motivation for completing the study might have been greater for their population of mandated college students.

Attrition data have been mainly coming from trials. Compared with the dropout and nonusage attrition rates in effectiveness trials of Web-based interventions, attrition rates in open-access interventions are higher [14]. This might be due to the use of participant-retention strategies in trials and to the characteristics of trial participants (eg, motivated participants). The study of Linke et al [18], with a follow-up and usage completion rate of 16.5%, involved a cohort study with 10,000 users of a free, Web-based, 6-week intervention. They used a strict definition of attrition, as only registrants who completed the whole 6-week program and the final assessment were considered to be completers. In comparison, Cunningham et al [11] noted that, despite their excellent 92% follow-up completion rate at 6 months, 35 of the 92 participants in the intervention condition (38%) never accessed the intervention. Riper and colleagues investigated their self-help intervention (Drinking Less) in a randomized controlled trial (RCT) and an open-access sample. They reported a 54% follow-up completion rate for the 6-month follow-up in the RCT intervention group, and 45% of the baseline participants actually made use of the intervention [13]. In their open-access sample, they found a follow-up completion rate of 40.5% but 12% of participants never using the program, 60% using it once or a few times, and 28% using the intervention more than a few times [22]. The study examples above illustrate that providing access to an intervention does not guarantee that participants use it.

The high percentages of nonusage attrition lead to the question of whether Web-based alcohol treatment might work more effectively for some people than for others. Exploring the variables that make individuals more vulnerable to not completing treatment may help us to identify target groups and develop strategies to address the nonusage attrition problem. We examined three types of variables that were associated with nonusage or dropout attrition: sociodemographic variables, drinking behavior, and psychological variables. Those factors have been investigated in several online alcohol intervention studies. Although most studies found no differences in baseline variables between completers and noncompleters [5,8-11,13,17,21], other studies did find support for baseline differences. Sociodemographic variables found to be positively associated with intervention and follow-up completion were being female [12,18], married or living with a partner [18,22], and without children [18]. Riper and colleagues also found that follow-up noncompleters were more likely to be above the median age of 47 years [22]. Chiauzzi et al [6] found that study site (2 out of 5 universities) was a predictor of follow-up noncompletion. Regarding baseline drinking-behavior variables, intervention completers showed less risk of alcohol dependency and harm from alcohol [18], and consumed fewer units a week and per occasion than noncompleters [20]. Additionally, psychological predictor variables were found in two studies. Chiauzzi et al [6] found baseline stage of readiness for change (contemplation) to be a predictor of dropout attrition, and Postel et al [16] found that intervention completers had a higher baseline score on treatment readiness. It could be suggested that

XSL•FO
RenderX

the results concerning the differences between completers and noncompleters are frequently ambiguous and are often found in only a single study. This might be the result of the differences in target groups and intervention characteristics. In line with this, Murray and colleagues [24] emphasized that it is important to adjust boosting strategies to the particular target population of the Web-based intervention. Whereas studies of online weight-loss programs, for example, have successfully boosted follow-up rates by using postal and telephone reminders for participants who did not respond to email reminders, Murray et al [24] received only 3% additional responses from their population of hazardous drinkers after an extensive additional follow-up using postal reminders and phone calls. It would be interesting to further investigate why such a strategy is working in one population but not in another one.

None of the Web-based alcohol intervention studies formally examined the reasons for noncompletion. Although most studies report the rates of nonusage or dropout attrition, they do not report the reasons for attrition. However, in our recently conducted RCT we examined the reasons for not completing treatment [16]. The Dutch Web-based treatment program (alcoholdebaas.nl) has been shown to be effective for problem drinkers in reducing their alcohol consumption and improving health status, yielding a large effect size at posttreatment [16]. The attrition rate in our Web-based treatment group (n = 42) was high at 54%. As we used a linear model for the treatment program with technically integrated assessment points, nonusage attrition automatically meant dropout attrition. Questionnaires could be sent to respondents only when all previous assignments were completed. Therefore, attrition was defined as not completing all 12 sessions of the Web-based intervention: 9 assignments and 3 assessments. We investigated reasons for noncompletion by sending an online questionnaire to all noncompleters. As described previously [16], the results showed that the main reasons for noncompletion in the Web-based treatment group were personal reasons unrelated to the Web-based treatment program, discomfort with the treatment protocol, and satisfaction with the positive results achieved to date. The present paper includes much more data regarding attrition in Web-based treatment for problem drinkers. We added the attrition data of the delayed control group and of a nontrial sample, and we conducted prediction analyses on pretreatment predictors of treatment completion. We also conducted qualitative analyses to get more insight into the reasons for dropout and participants' suggestions for how to enhance the number of treatment completers.

The first aim of this study was to examine attrition prevalence and pretreatment predictors of attrition in a cohort of open-access users of the Web-based treatment program. The second aim was to further explore attrition data from our RCT. We investigated the prevalence of attrition, the reasons for noncompletion, pretreatment predictors of attrition, and participants' suggestions for how to enhance treatment completion. Accordingly, the present study allowed us to compare the attrition data of both samples: a trial and an open-access group of users.

Methods

Study Design and Participants

The real-world sample consisted of all open-access users of the Web-based alcohol treatment program in 2009 (n = 885). The only inclusion criterion for open-access users was a minimum age of 18 years. All data entered by participants were stored in the Web-based application. We could identify who accessed the Web-based treatment program and who did not, the duration of participation for treatment completers, and the number of completed sessions in case of noncompletion. Participants who dropped out were not assessed about their situation at that time; because of the feasibility nature of the open-access study and the linear design it was not possible to send questionnaires to nonresponders through the application.

We conducted secondary analyses of our RCT: an open trial with participants randomly assigned to either the Web-based treatment group or to the waiting list control group [16]. The study protocol was approved by the independent medical ethics board METiGG (reference number NL20742.097.07) and registered at www.controlled-trials.com (ISRCTN39104853). In brief, we recruited Dutch-speaking problem drinkers in the general population aged ≥18 years. Problem drinking was defined as drinking currently at least 15 units (of 10 grams of ethanol) a week for women and 22 units a week for men. We excluded participants treated for problem drinking in the preceding year and participants with psychiatric treatment in the past 6 months or those currently with a psychiatric disorder. Of the problem drinkers screened (n=169), 156 were found to be eligible for the study, and they were randomly assigned to either the Web-based treatment group or to the waiting list control group. As the control group received the intervention immediately after the experimental group completed treatment, we merged the data from both groups for the present study. Participants received the e-therapy intervention free of charge. We did not provide any kind of incentive for study participation.

Intervention

The Web-based alcohol treatment consisted of a structured, 2-part, online treatment program in which the participant and the therapist communicated asynchronously, via the Internet only. The method underlying the program was based on the principles of cognitive behavior therapy [25] and motivational interviewing [26]. Part 1 of the program consisted of 2 assessments and 4 assignments and focused on the analysis of the participants' drinking habits. Part 2 focused on behavioral change and included 5 assignments and 1 final assessment. The average duration of the total treatment program was 3 months, with one or two therapist contacts per week and daily self-reporting of alcohol intake during the whole program. The 12 treatment sessions were identical for RCT and open-access users, except for the 3 assessments being more extensive for RCT participants.

Outcome Measures

Participants' pretreatment characteristics were derived online from the baseline self-report questionnaire, for RCT as well as for open-access participants. Weekly alcohol consumption was

XSL•FO
RenderX

assessed by a 7-day retrospective drinking diary, including a question about atypical drinking [27]. Type and severity of substance dependence were assessed by the Substance Abuse Module of the Composite International Diagnostic Interview [28]. The 28-item General Health Questionnaire (GHQ-28) and the Maudsley Addiction Profile-Health Symptom Scale (MAP-HSS) were used to assess health status [29,30]. The 21-item Depression Anxiety Stress Scale (DASS-21) was used to measure the three related negative emotional states of depression, anxiety, and stress [31]. To measure the quality of life, the EQ-5D was used [32]. Initial treatment motivation was measured with the TCU Motivation for Treatment (MfT) scale [33], and participants' readiness to change their drinking behavior was measured with the Dutch version of the Readiness to Change Questionnaire [34]. For open-access participants the questionnaires were less extensive, as the GHQ-28 and MAP-HSS were left out.

The outcome measure of the logistic regression analysis was completion of the Web-based alcohol treatment program; this was defined as completion of all 12 treatment sessions: 9 assignments and 3 assessments. Because of the linear design of the treatment program it was impossible for participants to skip parts of the intervention; therefore, the point at which they stopped using the program indicates exactly how much treatment participants received. In our study nonusage attrition automatically meant dropout attrition and we will therefore just use the term attrition.

In order to gain insight into the motives of participants to stop using the Web-based treatment program, noncompleters in the RCT group received an email with a link to an additional online questionnaire consisting mainly of open questions concerning their perception of the program, reasons for discontinuation, and suggestions to improve the intervention and enhance treatment completion. If participants did not complete this questionnaire, they were contacted by telephone to remind them to complete the questionnaire either online or alternatively by phone.

Statistical Analysis

Chi-square and *t* tests were used to assess whether there were baseline differences between completers and noncompleters. Multivariate logistic regression analysis was performed with treatment completion as the dependent variable. Predictor variables with $P < .10$ in the univariate analyses were entered in a full multivariate model. Subsequently, nonsignificant variables were removed, one by one, until –2 log likelihood deteriorated significantly. Goodness of fit of the model was determined by the Hosmer-Lemeshow test, and the Nagelkerke R^2 was used for the pseudo proportion of variance. Three regression analyses were performed concerning (1) the RCT sample, (2) the open-access sample including treatment readiness variable, and (3) the open-access sample without treatment readiness variable. Because treatment readiness was measured after part 1 in the open-access sample, we had a lot of missing data for this variable (n = 355). We therefore performed two regression analyses for the open-access sample, one including treatment readiness (and as a consequence only the noncompleters from part 2) and one without this variable (all noncompleters). The predictor variables for the RCT sample were age, gender, work, education level, baseline alcohol consumption, prior alcohol treatment, prior mental health treatment, readiness to change contemplation, and action score, DASS-21 total score, and the MfT questionnaire scores for desire for help and treatment readiness. For the open-access sample, the DASS-21 scores were not available and therefore left out of the regression analysis. All statistical tests were 2-sided, with $P \leq .05$ considered to be significant, and performed using SPSS for Windows 17.0 (IBM Corporation, Somers, NY, USA).

Reasons for nonusage attrition were independently assessed by the first and third author (qualitative study). The agreement level between both authors was 87%, which was considered acceptable. If the two authors did not agree, the topic was discussed in order to reach agreement. Participants' responses to open questions were analyzed using thematic analysis. The first author carefully searched through the data to identify and code all features concerning participants' reasons for not completing the treatment program. After collating relevant data with each code, related patterns were combined into themes. After refining and defining the themes, a brief description of each theme was formulated related to the research questions of the study.

Results

Participant Characteristics

Of the 885 registrants for the open-access version in 2009, 105 never started using the Web-based alcohol treatment program by doing the first assignment, sending a message to their therapist, or logging into the daily alcohol diary. Of the 780 participants who started the open-access version, 54.0% (n = 421) were women, 49.6% (n = 387) had a higher education level, and 69.0% (n = 538) were employed. Age ranged from 20 to 78 years, with an average of 45.7 years (Table 1). A total of 689 participants reported alcohol dependence (88.3%), but many (n = 554, 71.0%) had never received professional help for their drinking problem. The mean weekly alcohol consumption was 42.7 standard units a week: 49.1 for men and 37.3 for women.

XSL•FO
RenderX

Table 1. Characteristics of participants in the randomized controlled trial (RCT) and open-access group

Variable	RCT participants (n = 144)		Open-access participants (n = 780)	
	n	%	n	%
Female	83	58	421	54.0
Higher education	84	58	387	49.6
Employed	117	81.3	538	69.0
DSM-IV[a] diagnosis				
Alcohol dependence	120	83.3	684	87.7
Alcohol abuse	14	10	42	5
No dependence or abuse	10	7	54	7
Prior treatment for alcohol abuse	22	15	226	29.0
Prior treatment mental health problems	72	50	455	58.3
Problem drinking[b]	144	100	689	88.3
	Mean	SD	Mean	SD
Age (years)	45.8	9.7	45.7	10.8
Weekly alcohol consumption (standard units/week)				
Men	49.8	26.9	49.1	30.1
Women	32.6	14.6	37.3	22.9
GHQ-28 score[c]	52.6	11.9	NA[d]	NA[d]
MAP-HSS score[e]	19.8	6.2	NA[d]	NA[d]
DASS-21[f]				
Depression score	8.7	8.4	NA[d]	NA[d]
Anxiety score	5.9	5.9	NA[d]	NA[d]
Stress score	12.5	8.2	NA[d]	NA[d]
RCQ[g]				
Precontemplation	12.1	1.3	12.3	1.6
Contemplation	17.1	2.1	17.1	2.3
Action	12.4	3.5	13.3	3.3
MfT[h]				
Treatment Readiness	4.0	0.5	4.1	0.4
Desire for Help	3.9	0.7	3.9	0.6

[a]*Diagnostic and Statistical Manual of Mental Disorders*, 4th revision.

[b] Drinking >21 (men) or >14 (women) mean units per week.

[c] 28-item General Health Questionnaire.

[d] Not applicable.

[e] Maudsley Addiction Profile-Health Symptom Scale.

[f] 21-item Depression Anxiety Stress Scale.

[g] Readiness to Change Questionnaire.

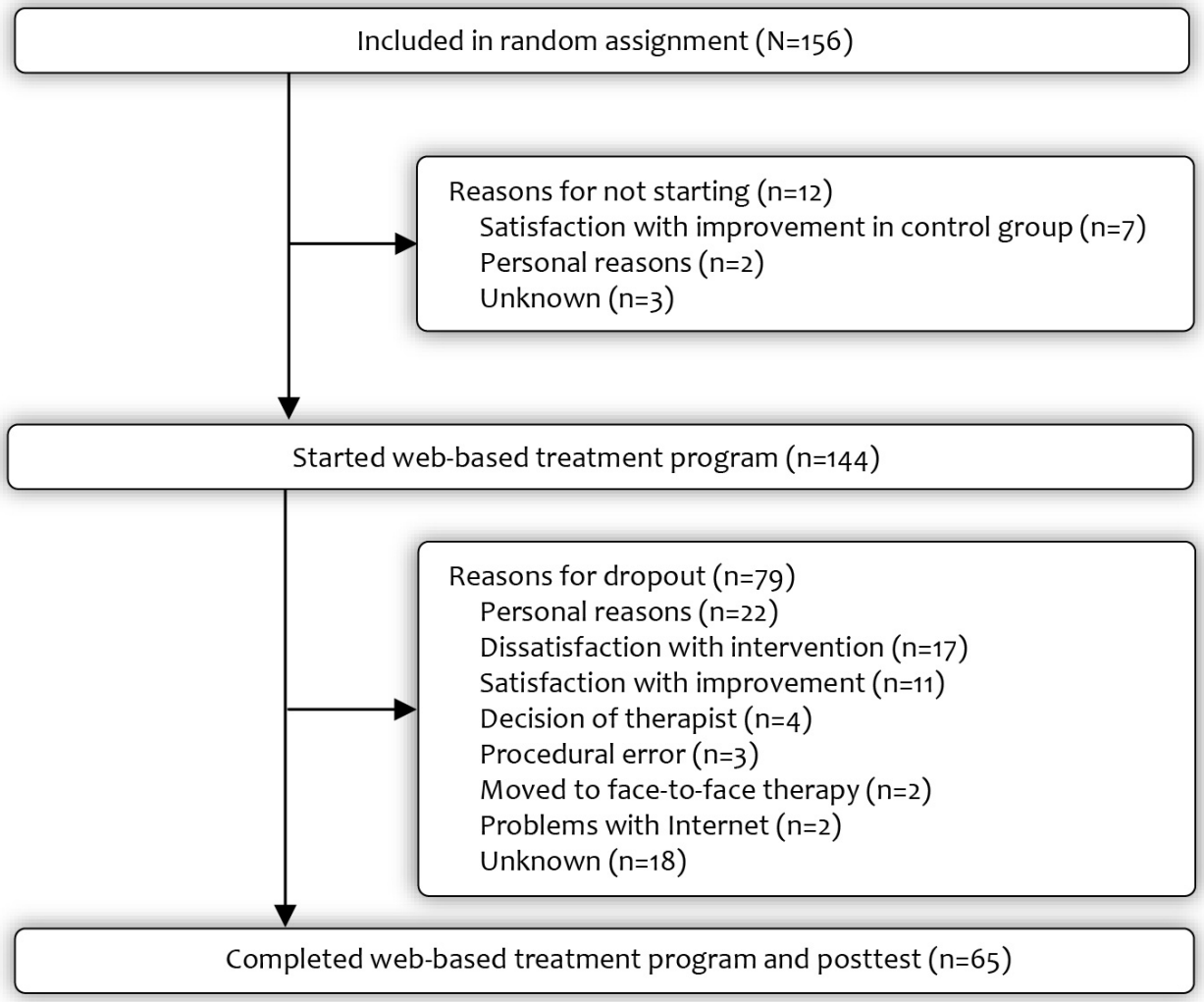
[h] TCU Motivation for Treatment scale.

Figure 1 shows the participant flow of the total RCT sample (n = 144) along with reasons for not starting (n = 12). Pretreatment characteristics of the 144 RCT participants who started the Web-based treatment program are presented in Table 1. Of these participants, 58% (n = 83) were women, 58% (n = 84) had a higher education level, and 81.3% (n = 117) were employed. Ages ranged from 22 to 66 years, with an average of 45.8 years, and 120 participants reported dependence (83.3%). The majority

XSL•FO
RenderX

(n = 122, 84.7%) had never received professional help for their drinking problem. The mean weekly alcohol consumption was 39.9 standard units a week: 49.8 for men and 32.6 for women.

Figure 1. Flow of participants in the randomized controlled trial.



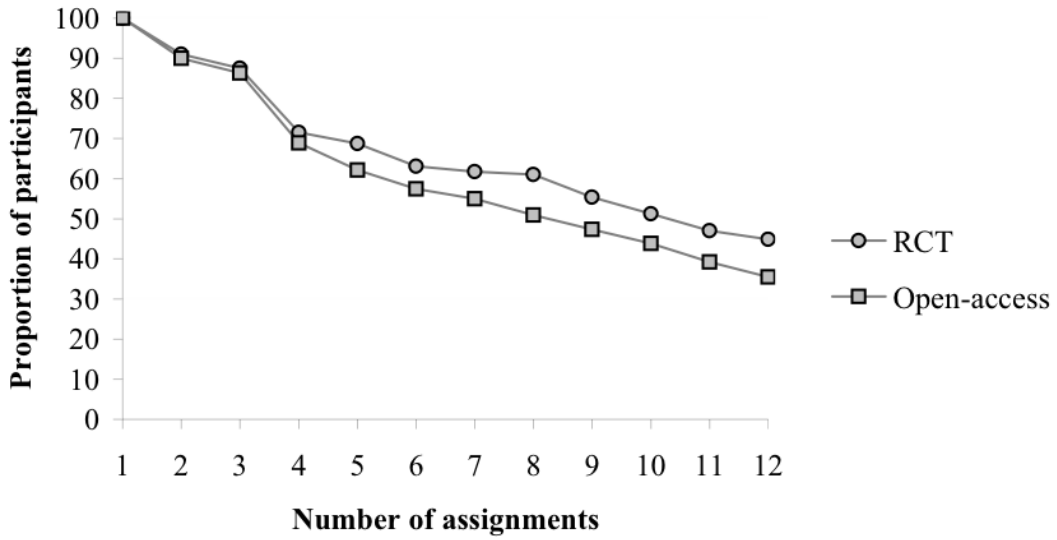
Attrition Pattern

Of the 780 open-access participants, 65.0% were noncompleters. Treatment completers (n = 273, 35.0%) completed all 12 treatment sessions and noncompleters (n = 507, 65.0%), an average of 4.8 (SD 3.1) sessions. Of the 144 RCT participants, 55% were noncompleters. Treatment completers (n = 65, 45%) completed all 12 treatment sessions and noncompleters (n = 79, 55%), an average of 4.8 (SD 3.1) sessions. The open-access and RCT group differed significantly in the percentage of treatment completers ($\chi^2_1 = 5.4$; $P = .02$). Participants in the RCT sample were 1.29 (95% confidence interval [CI] 1.05–1.58) times more likely to complete treatment.

Participants completed the sessions in the order that they were presented. The average duration of treatment to completion was 16.1 weeks in the RCT sample and 17.1 weeks in the open-access sample. Figure 2 shows the attrition curves of both groups. Participants dropped out during all stages of treatment. However, the biggest loss was found after the third session, possibly as a result of the daily drinking diary. In this session, participants were asked to register daily amounts of alcohol consumption for the whole treatment duration.

XSL•FO
RenderX

Figure 2. Attrition curve: proportion of participants by number of assignments in the randomized controlled trial (RCT) and open-access group.



Predictors of Treatment Completion

We found only one significant difference between completers and noncompleters in the RCT sample. The mean score on the Treatment Readiness subscale of the MfT was higher for completers (mean 4.13) than for noncompleters (mean 3.97), $t_{1,142}$ = –2.00, *P* = .047. There were no other significant differences between the groups on any of the variables presented in Table 1. Logistic regression analysis revealed a statistically significant contribution of treatment readiness score. The regression equation showed a negative predicted value of 70% and a positive predictive value of 53%, with a cut-off probability of the model of 0.4. The Nagelkerke *R2* was .04, and the regression model showed sufficient goodness of fit (χ^2_1 = 10.7, *P* = .22). The area under the receiver operating characteristic (ROC) curve revealed a model discrimination value of 0.60 (95% CI 0.51–0.70). The odds ratio indicated that if the treatment readiness score increases by 1 point (range 1–5), the odds of completion increase with 2.1. A score of 3 gives a chance of completion of 27%, a score of 4 a chance of 44%, and a score of 5 a 63% chance.

We found seven significant differences between completers and noncompleters in the open-access sample: age, gender, education level, baseline alcohol consumption, prior mental health treatment, treatment readiness, and readiness to change action score. The differences are shown in Table 2. Multivariate logistic regression analysis with treatment readiness included (n = 425) revealed a statistically significant independent contribution of age, baseline alcohol consumption, and treatment readiness. Predicted probabilities of the model of x, y, and z led to a specificity of 89% with a sensitivity of 25%, a specificity of 84% with a sensitivity of 33%, and a specificity of 78% with a sensitivity of 40%, respectively. The Nagelkerke *R2* was .09, and the regression model showed sufficient goodness of fit (χ^2_1 = 11.7, *P* = .17). The area under the ROC curve revealed a discrimination of the model of 0.64 (95% CI 0.59–0.70). When the treatment readiness score increases by 1 point (range 1–5), the odds of completion increase 2.1-fold. If age increases by 5 years, the odds of completion increase 1.12-fold, and if baseline alcohol consumption increases by 10 standard units a week, the odds of completion decrease 0.87-fold.

XSL•FO
RenderX

Table 2. Differences between open-access completers and noncompleters

Variable	Completers (n = 273)		Noncompleters (n = 507)		Test result		
	n	%	n	%	χ^2	df	*P* value
Female	170	62.3	251	48.5	11.6	1	<.001
Higher education	163	59.7	224	44.2	17.1	1	<.001
Prior mental health treatment	175	64.1	280	55.2	5.8	1	.02
	Mean	SD	Mean	SD	*t*	df	*P* value
Age (years)	47.8	10.4	44.5	10.9	–4.14	1,778	<.001
Baseline alcohol consumption (standard units/week)	37.4	24.0	45.6	28.2	4.05	1,778	<.001
MfT[a] Treatment Readiness	4.1	0.4	4.0	0.4	–3.30	1,423	.001
RCQ[b] action score	13.8	3.3	13.0	3.3	–3.43	1,778	<.001

[a] TCU Motivation for Treatment scale.

[b] Readiness to Change Questionnaire.

Multivariate logistic regression analysis without treatment readiness (n = 780) revealed a statistically significant contribution of age, gender, education level, baseline alcohol consumption, and readiness to change action score. Predicted probabilities of the model of x, y, and z led to a specificity of 85% with a sensitivity of 25%, a specificity of 80% with a sensitivity of 35%, and a specificity of 75% with a sensitivity of 43%, respectively. The Nagelkerke *R2* was .10, and the regression model showed sufficient goodness of fit ($\chi^2_1 = 7.1$, $P = .53$). The area under the ROC curve revealed a discrimination of the model of 0.63 (95% CI 0.59–0.67). The odds of treatment completion was 1.70-fold increased for women compared with men, and 1.79-fold increased for people with higher education compared with less-educated people. The odds ratios further indicated that if age increases by 5 years, the odds of completion increase 1.13-fold, and if baseline alcohol consumption increases by 10 standard units, the odds of completion decrease 0.93-fold. If the readiness to change action score increases by 1 point (range 4–20), the odds of completion increase 2.1-fold.

Early Versus Late Noncompleters

We divided noncompleters into early and late noncompleters to determine whether the two groups differed. We considered noncompleters who completed a maximum of 3 assignments to be early noncompleters and those who completed at least 4 assignments to be late noncompleters. We found no differences between both groups in the RCT sample (n = 144). However, in the open-access sample (n = 780) we found that, compared with those who completed fewer assignments, more noncompleters who completed at least 4 assignments had a high level of education (128/221, 57.9% vs 93/221, 42%, $\chi^2_1 = 6.1$, $P = .01$), had received prior mental health treatment (162/276, 58.7% vs 114/276, 41.3%, $\chi^2_1 = 12.0$, $P < .001$), and had a lower baseline alcohol consumption (43.2 vs 48.3 standard units a week, $t_{501} = 2.01$, $P = .045$).

Reasons for Noncompletion

Figure 1 shows the reasons for noncompletion (n = 79). Self-reported reasons for not completing treatment were collected only in the RCT sample, and were obtained from 61 of 79 participants (77%). We were not able to contact 18 participants because of nonresponse or an invalid phone number. The most common reason for not completing treatment consisted of personal reasons unrelated to the Web-based intervention (n = 22), followed by dissatisfaction with the intervention (n = 17), and satisfaction with the improvement in their condition (n = 11). On four occasions the therapist decided to terminate the treatment, because of insufficient response or information (n = 3) or due to an inability to set a realistic drinking goal (n = 1). Unfortunately, in three cases we had procedural problems during the trial, and those participants could not continue. Additionally, 2 participants moved on to face-to-face treatment and 2 participants experienced problems with the Internet during treatment participation.

Personal Reasons

A diversity of personal reasons were given as reason for noncompletion (n = 22), including being too busy with work, a seriously ill family member or bereavement, other priorities, a hospitalization, no Internet access, or moving house.

Dissatisfaction With Intervention

Participants who identified the Web-based alcohol intervention itself as a reason for discontinuation (n = 17) most commonly indicated that the program was too time consuming or too demanding. Some participants reported that the program could not meet their personal needs.

Improvement in Condition

Several participants reported that they no longer felt the need to continue the program, because of the progress they made (n = 11). They gained from the intervention what they needed and felt in control of their drinking behavior.

XSL•FO
RenderX

Other Reasons

For 2 participants the Web-based treatment program was only the first step in working on behavioral change, and they continued treatment in a face-to-face setting. Of the persons whose formal reason for dropout is unknown (n = 18), the messages in their personal records provide some information. Participants mentioned several times that working on their alcohol problem was quite confrontational and overwhelmed them too much. Some participants also reported more or less lack of motivation.

Suggestions to Enhance Treatment Completion

Several RCT participants gave suggestions as how to improve the Web-based treatment program. One of the suggestions was sending an email message to participants to notify them that they had received a new message or assignment from their therapist. This it was felt would act as a reminder and prevent unnecessary logging into the application. Another suggestion was to allow more flexibility in the treatment protocol, with the possibility of skipping sessions when required—for example, the possibility to start immediately with the goal-setting assignment or no longer mandating daily registration. In its current form it was not possible to move on to the next assignment without completing the previous one. Some participants also mentioned the need for additional contact: the choice to contact their therapist by phone or face-to-face and the chance to get in touch with fellow participants, with the suggestion to link each participant to his or her own buddy. Some participants made suggestions for improving the usability of the Web-based treatment program, including the speed of the intervention, layout characteristics, and button functions.

Discussion

Main Findings

The aim of this study was to explore the attrition data of an open-access and an RCT sample of a Web-based treatment program for problem drinkers. The study demonstrated high prevalence of attrition in both samples, with 10% less treatment completers in the open-access sample. Participants' readiness for treatment, gender, education level, age, baseline alcohol consumption, and readiness to change score were shown to predict treatment completion. The key reasons for noncompletion were personal reasons, dissatisfaction with the intervention, and satisfaction with their own improvement. The main suggestions for boosting strategies involved email notification and more flexibility in the intervention.

Attrition

Attrition was high in both samples. Although our attrition rates of 65% in the open-access sample and 55% in the RCT sample are in line with those found in other Web-based alcohol intervention studies [12,13,22], the majority of alcohol intervention studies found lower attrition rates [4-6,8-11,17,21]. However, comparing attrition rates alone does not make sense. A clear description of the study characteristics together with nonusage and follow-up attrition is necessary to interpret attrition data properly. Our attrition rates need to be seen in the light of a strict definition of treatment completion including assessment completion, active usage of the intervention, a high intensity of the treatment program, and paying no incentive to participants. In comparison, Linke et al [18] used a similar definition of attrition in their cohort sample of the brief intervention Down Your Drink and found a completion rate of 16.5%. To the best of our knowledge, no online alcohol intervention studies have been published concerning comparable guided treatment with intensive therapist contact. We therefore can only compare our attrition rates with those of more or less intensive online alcohol self-help interventions. Although there is some evidence from computer-aided psychological treatment programs that participants receiving extra therapist contact (eg, phone support) drop out less often, no studies have explored the influence of therapist contact on dropout from Web-based treatments for psychological disorders [35]. Further investigation of the impact of therapist contact on attrition from online alcohol interventions is needed.

The variety of nonusage and dropout attrition rates in Web-based alcohol interventions is relatively similar to that found in Web-based treatments for psychological disorders, ranging from 2% to 83% [35]. A higher number of noncompleters in our open-access sample is consistent with earlier findings [14]. The fact that RCT participants were 1.5 times more likely to complete treatment might be the result of a selection bias, because of the prescreening of trial participants and the exclusion criteria. It leads to the suggestion that it might be wise to always link some kind of research to a Web-based intervention and to emphasize the importance of it at the start. Realizing that you are cooperating in a research project, for example to improve the intervention, can perhaps be inspiring. We acknowledge that it is important to find a good balance between what is needed for attrition purposes and what is considered to be ethically appropriate. Finding the right tone seems to be important. Further research needs to investigate whether this strategy will be effective in reducing the number of dropouts, and whether this works for participants and for therapists. What is the impact of this for participants? Do therapists change the treatment or the communication with participants if they know that the data will be used for research purposes? Are therapists extra motivated to increase adherence to the treatment protocol?

In both study samples, the pattern of nonusage attrition was steady throughout the intervention period. This means that both groups showed the same trend of attrition; at each treatment session participants dropped out. The number of dropouts gradually decreased, regardless of whether participants participated in the RCT or in the open-access intervention. Although the gradual decrease is in contrast with the suggestion of Eysenbach [15] that, in the final stage of an intervention, a hardcore group of users remain who will continue using the intervention, it is identical to the attrition pattern found by Neve and colleagues [36] in their 12-week, Web-based weight-loss program.

The percentage of dropouts seems to be the highest after session 3, concerning the daily drinking diary assignment. A possible explanation might be the intensity of this assignment, as participants have to register their alcohol consumption every day. This might be quite confrontational and participants might

XSL•FO
RenderX

also feel uncomfortable or annoyed by daily registration of their drinking amount.

The differences we found between early and late noncompleters prove that noncompleters who completed at least 4 assignments were more similar to treatment completers than they were to those who completed fewer than 4 assignments.

Predictors of Completion

The only statistically significant predictor of treatment completion in the RCT sample was a higher treatment readiness score, measured by the Treatment Readiness subscale of the TCU MfT questionnaire. In the open-access sample, higher treatment readiness also was a significant predictor, as were higher age and lower baseline alcohol consumption when the treatment readiness variable was included (n = 425). In the open-access sample without the treatment readiness variable (n = 780), the statistically significant predictors were higher age, female gender, higher education level, lower baseline alcohol consumption, and higher readiness to change action score. Other factors were found to have no predictive value.

Based on our different findings in the three subsamples and in line with an analysis of the literature by Melville and colleagues [35], we have to conclude that the current evidence for predictors of attrition is ambiguous. Two other Web-based alcohol intervention studies previously found that study completers consumed less alcohol at baseline [20,22]. Earlier studies by Bewick et al [12] and Lange et al [37] also found that more men than women were noncompleters, although Riper et al [22] did not find a significant association between gender and dropout. Male gender was also found to be associated with noncompliance in face-to-face addiction treatment [19]. A higher education level as a predictor of completion was not confirmed by three studies that explored the influence of education level on dropout from Web-based interventions; they did not find a significant association [22,37,38]. However, the association between compliance and higher education level was confirmed in face-to-face addiction treatment [19]. With regard to age, previous evidence was contrary to our findings. Riper et al [22] found that noncompleters were more likely to be above the median age of 47 years, whereas we found that noncompleters were younger than completers. Previous Web-based intervention studies also did not confirm the differences in treatment readiness between completers and noncompleters and found no predictive value for readiness to change [12]. But lower intention to comply with treatment and weaker initial treatment motivation were found to be associated with noncompliance in face-to-face addiction treatment [19]. The relationship between the baseline variables and dropout might also be mediated by other variables. Older participants or more highly educated participants might, for example, use the Internet in a different manner from younger or less-educated participants. Women probably experience more support from their relatives, which might stimulate continuation of treatment. And participants with lower baseline alcohol consumption may have more confidence in their own effectiveness. It would be interesting to further investigate the relationship between baseline variables and dropout. Overall, our findings also raise the question of how useful this kind of prediction research is. Because of the considerable variation in findings, we would on the one hand suggest that further research is needed to confirm whether the same predictors exist in different Web-based alcohol interventions, but on the other hand we would also suggest not focusing too much on baseline predictors of online treatment completion. It might be more effective to focus on the therapist side and the effects of boosting strategies in online interventions. The clinical implications of this study can therefore only be given with caution. It would be interesting to investigate whether increasing treatment readiness and readiness to change immediately from the start of treatment would decrease the number of noncompleters. Additionally it might be interesting to find out whether it matters how fast participants reduce their alcohol consumption or become abstinent after the start of the treatment program. Another question could be whether the pace at which participants experience a positive relationship with their therapist also has an effect on treatment completion.

Reasons for Noncompletion

In addition to the quantitative data of the RCT and open-access sample, the qualitative data provided more insight into the reasons for noncompletion and the possibilities to reduce potential loss. The present more extensive findings confirm the earlier findings on dropout from our RCT study and, as discussed before [16], most reasons for noncompletion are in line with the potential factors for attrition as described in the law of attrition by Eysenbach [15], except for improvement in condition. Some participants significantly improved after just a few treatment sessions, and they were convinced that no additional sessions were needed. This confirms Christensen and Mackinnon's statement that low usage and dropout do not necessarily coincide with failure [39]. Participants who do not complete the treatment program or follow-up assessments may still derive much benefit from the Web-based intervention. Continuous and frequent measurement, such as with diary surveys, can provide the necessary data [40]. Although a disadvantage of diary surveys is that the respondents themselves are responsible for completion, a Web-based intervention has the potential to easily prompt users by automatically sending reminders, motivational messages, or incentives. We also suggest investing in easy referral from Web-based treatment to face-to-face treatment with the possibility of integrated treatment (Web-based and face-to-face). Participants as well as their therapists expressed interest in this kind of integrated care. Professionals at the International Network on Brief Interventions for Alcohol Problems conference also expressed interest in this possibility [41].

Boosting Strategies

Boosting strategies are desirable to maximize the number of treatment completers in trial settings as well as in open-access interventions. Participants themselves suggested sending email reminders as an additional supportive resource. The use of push reminders, such as phone calls, postcards, and email messages, previously has shown improved treatment completion rates [42,43]. Although participants already received therapists' messages in the Web-based application, they preferred receiving reminders in their private email account in order to be constantly reminded of their participation and to prevent unnecessary

XSL•FO
RenderX

logging into the application. Participants also suggested more flexibility in the Web-based treatment program. The most frequently mentioned response was that daily alcohol registration was somewhat annoying to participants. This might explain the more pronounced loss of participants (16%–17%) after the third session, as this assignment requested starting with daily alcohol registration. Another suggestion was to better adapt the pace of the treatment to the needs of the individual participant and not being too rigid in terms of the fixed treatment duration. Interestingly, none of the participants suggested incentives as a useful boosting strategy, possibly because they thought this was embarrassing to suggest. Contingency management interventions have been shown to increase desired behavior by offering valuable reinforcements contingent on behavioral change [44]. It would be an interesting direction for future research to apply the contingency management principles in Internet interventions and to investigate their effectiveness.

Methodological Considerations

This study has several limitations that are important to acknowledge. Due to the technical structure of our intervention, noncompletion included not just stopping using the intervention but also no longer receiving posttest and follow-up assessments. The therapists and participants could not move on to the next assignment or questionnaire without completing the previous one. We chose this linear model because of the protocolled treatment and the preference for completing treatment steps in strict order, to ensure best quality and that the questionnaires would be completed. However, a consequence that we have not sufficiently taken into account is that nonusage attrition also meant study attrition and that we unfortunately never obtained a lot of data from noncompleters. This is definitely not desirable and needs to be changed in future studies. One of the consequences is that we did not have data available to compare treatment outcome of completers versus noncompleters. Although our qualitative data indicated that completers had better treatment results, this assumption can be confirmed only with quantitative data. As far as we know, no previous online alcohol intervention study has investigated the difference in treatment outcome between completers and noncompleters. We therefore recommend investigating the impact of compliance on treatment outcome in future studies.

We also decided not to use push factors in our RCT to keep the trial setting as natural as possible. However, it is possible that, if we had used push factors, we could have raised the response rate to generate a more complete dataset.

Another limitation is that only baseline characteristics were considered as potential predictors of treatment completion. It is possible that other factors such as forum use or the therapeutic relationship also influenced attrition rate. However, we aimed to determine at baseline which participants would complete the whole treatment program. We were also limited to the baseline characteristics we measured and therefore not able to include some of the variables previously found to have predictive value.

Both study samples consisted largely of adults in their mid-40s. This can partly be explained because our samples consisted of problem drinkers from the general public. And although we previously found that the average age of face-to-face clients was slightly lower, face-to-face clients also have a mean age of around 43 years [1,45]. It often takes a long time before people experience excessive alcohol consumption as a problem. The physical and psychological damage will only be felt over time. People in their mid-40s often take responsibility for their own health and are looking for a healthier lifestyle, including drinking less. Web-based treatment is a pleasant option for them, because of the privacy and easy access to online help. Although they are an important target group for our intervention, it remains a challenge to reach younger and older problem drinkers via the Internet as well. Future research should focus on how these groups can be reached.

Future Directions and Implications

Nowadays, the challenge of Web-based alcohol treatment programs no longer seems to be their effectiveness but keeping participants involved until the end of the treatment program. Our study provided some points that therapists might focus on, including helping participants to be ready for treatment and for change. We should also investigate the effect of starting immediately with reduction of alcohol consumption. Boosting strategies such as email notification and more flexibility in the intervention might also help to improve adherence. Further research should investigate whether those changes lead to decreased attrition rates in Web-based interventions. If we can succeed in improving attrition rates, we assume that the success of Web-based alcohol interventions will further improve and, as a consequence, they will have a greater public health impact.

Acknowledgments

This study was funded by Tactus Addiction Treatment and the Nijmegen Institute of Scientist-Practitioners in Addiction.

Conflicts of Interest

None declared

References

1. Postel MG, de Jong CAJ, de Haan HA. Does e-therapy for problem drinking reach hidden populations? Am J Psychiatry 2005 Dec;162(12):2393. [doi: 10.1176/appi.ajp.162.12.2393] [Medline: 16330613]

XSL•FO
RenderX

2. 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]
3. Bewick BM, Trusler K, Barkham M, Hill AJ, Cahill J, Mulhern B. The effectiveness of web-based interventions designed to decrease alcohol consumption--a systematic review. Prev Med 2008 Jul;47(1):17-26. [doi: 10.1016/j.ypmed.2008.01.005] [Medline: 18302970]
4. Saitz R, Palfai TP, Freedner N, Winter MR, Macdonald A, Lu J, et al. Screening and brief intervention online for college students: the ihealth study. Alcohol Alcohol 2007 Feb;42(1):28-36 [FREE Full text] [doi: 10.1093/alcalc/agl092] [Medline: 17130139]
5. 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. [doi: 10.1111/j.1360-0443.2004.00847.x] [Medline: 15500594]
6. Chiauzzi E, Green TC, Lord S, Thum C, Goldstein M. My student body: a high-risk drinking prevention web site for college students. J Am Coll Health 2005 Jun;53(6):263-274. [doi: 10.3200/JACH.53.6.263-274] [Medline: 15900990]
7. Moore MJ, Soderquist J, Werch C. Feasibility and efficacy of a binge drinking prevention intervention for college students delivered via the Internet versus postal mail. J Am Coll Health 2005 Aug;54(1):38-44. [doi: 10.3200/JACH.54.1.38-44] [Medline: 16050327]
8. Walters ST, Vader AM, Harris TR. A controlled trial of web-based feedback for heavy drinking college students. Prev Sci 2007 Mar;8(1):83-88. [doi: 10.1007/s11121-006-0059-9] [Medline: 17136461]
9. Doumas DM, Hannah E. Preventing high-risk drinking in youth in the workplace: a web-based normative feedback program. J Subst Abuse Treat 2008 Apr;34(3):263-271. [doi: 10.1016/j.jsat.2007.04.006] [Medline: 17600650]
10. Doumas DM, McKinley LL, Book P. Evaluation of two Web-based alcohol interventions for mandated college students. J Subst Abuse Treat 2009 Jan;36(1):65-74. [doi: 10.1016/j.jsat.2008.05.009] [Medline: 18657941]
11. Cunningham JA, Wild TC, Cordingley J, van Mierlo T, Humphreys K. A randomized controlled trial of an internet-based intervention for alcohol abusers. Addiction 2009 Dec;104(12):2023-2032. [doi: 10.1111/j.1360-0443.2009.02726.x] [Medline: 19922569]
12. Bewick BM, West R, Gill J, O'May F, Mulhern B, Barkham M, et al. Providing web-based feedback and social norms information to reduce student alcohol intake: a multisite investigation. J Med Internet Res 2010 Dec;12(5):e59 [FREE Full text] [doi: 10.2196/jmir.1461] [Medline: 21169171]
13. Riper H, Kramer J, Smit F, Conijn B, Schippers G, Cuijpers P. Web-based self-help for problem drinkers: a pragmatic randomized trial. Addiction 2008 Feb;103(2):218-227. [doi: 10.1111/j.1360-0443.2007.02063.x] [Medline: 18199300]
14. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
15. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
16. Postel MG, de Haan HA, ter Huurne ED, Becker ES, de Jong CAJ. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. J Med Internet Res 2010 Dec;12(4):e68 [FREE Full text] [doi: 10.2196/jmir.1642] [Medline: 21163776]
17. Finfgeld-Connett D, Madsen R. Web-based treatment of alcohol problems among rural women. J Psychosoc Nurs Ment Health Serv 2008 Sep;46(9):46-53. [Medline: 18823000]
18. Linke S, Murray E, Butler C, Wallace P. Internet-based interactive health intervention for the promotion of sensible drinking: patterns of use and potential impact on members of the general public. J Med Internet Res 2007 May;9(2):e10 [FREE Full text] [doi: 10.2196/jmir.9.2.e10] [Medline: 17513281]
19. de Weert-van Oene GH, Burger H, Grobbee DE, Schrijvers AJ. Identification of patients at high risk of noncompliance in substance dependence treatment. Eur Addict Res 2007;13(2):74-80. [doi: 10.1159/000097936] [Medline: 17356278]
20. Bewick BM, Trusler K, Mulhern B, Barkham M, Hill AJ. The feasibility and effectiveness of a web-based personalised feedback and social norms alcohol intervention in UK university students: a randomised control trial. Addict Behav 2008 Sep;33(9):1192-1198. [doi: 10.1016/j.addbeh.2008.05.002] [Medline: 18554819]
21. Kramer J, Riper H, Lemmers L, Conijn B, van Straten A, Smit F. Television-supported self-help for problem drinkers: a randomized pragmatic trial. Addict Behav 2009 May;34(5):451-457. [doi: 10.1016/j.addbeh.2008.12.015] [Medline: 19179014]
22. Riper H, Kramer J, Conijn B, Smit F, Schippers G, Cuijpers P. Translating effective web-based self-help for problem drinking into the real world. Alcohol Clin Exp Res 2009 Aug;33(8):1401-1408. [doi: 10.1111/j.1530-0277.2009.00970.x] [Medline: 19413646]
23. Riper H, Kramer J, Keuken M, Smit F, Schippers G, Cuijpers P. Predicting successful treatment outcome of web-based self-help for problem drinkers: secondary analysis from a randomized controlled trial. J Med Internet Res 2008 Nov;10(4):e46 [FREE Full text] [doi: 10.2196/jmir.1102] [Medline: 19033150]
24. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res 2009 Apr;11(2):e9 [FREE Full text] [doi: 10.2196/jmir.1052] [Medline: 19403465]

XSL•FO
RenderX

25. Hester RK, Miller WR, editors. Handbook of Alcoholism Treatment Approaches: Effective Alternatives. 2nd edition. Boston, MA: Allyn and Bacon; 1995.
26. Miller WR, Rollnick S. Motivational Interviewing: Preparing People for Change. 2nd edition. New York, NY: The Guilford Press; 2002.
27. Sobell MB, Sobell LC. Stepped care as a heuristic approach to the treatment of alcohol problems. J Consult Clin Psychol 2000 Aug;68(4):573-579. [Medline: 10965632]
28. Compton WM, Cottler LB, Dorsey KB, Spitznagel EL, Mager DE. Comparing assessments of DSM-IV substance dependence disorders using CIDI-SAM and SCAN. Drug Alcohol Depend 1996 Jul;41(3):179-187. [Medline: 8842630]
29. Marsden J, Gossop M, Stewart D, Best D, Farrell M, Lehmann P, et al. The Maudsley Addiction Profile (MAP): a brief instrument for assessing treatment outcome. Addiction 1998 Dec;93(12):1857-1867. [Medline: 9926574]
30. Vallejo MA, Jordán CM, Díaz MI, Comeche MI, Ortega J. Psychological assessment via the internet: a reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). J Med Internet Res 2007 Jan;9(1):e2 [FREE Full text] [doi: 10.2196/jmir.9.1.e2] [Medline: 17478411]
31. Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychol Assess 1998 Jun;10(2):176-181. [doi: 10.1037/1040-3590.10.2.176]
32. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Ned Tijdschr Geneeskd 2005 Jul 9;149(28):1574-1578. [Medline: 16038162]
33. De Weert-Van Oene GH, Schippers GM, De Jong CAJ, Schrijvers CA. Motivation for treatment in substance-dependent patients: psychometric evaluation of the TCU motivation for treatment scales. Eur Addict Res 2002 Jan;8(1):2-9. [Medline: 11818688]
34. Defuentes-Merillas L, Dejong CA, Schippers GM. Reliability and validity of the Dutch version of the Readiness to Change Questionnaire. Alcohol Alcohol 2002 Feb;37(1):93-99 [FREE Full text] [Medline: 11825864]
35. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
36. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. J Med Internet Res 2010 Dec;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]
37. Lange A, Rietdijk D, Hudcovicova M, van de Ven JP, Schrieken B, Emmelkamp PM. Interapy: a controlled randomized trial of the standardized treatment of posttraumatic stress through the internet. J Consult Clin Psychol 2003 Oct;71(5):901-909. [doi: 10.1037/0022-006X.71.5.901] [Medline: 14516238]
38. Ström L, Pettersson R, Andersson G. Internet-based treatment for insomnia: a controlled evaluation. J Consult Clin Psychol 2004 Feb;72(1):113-120. [doi: 10.1037/0022-006X.72.1.113] [Medline: 14756620]
39. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
40. Couper MP. Technology trends in survey data collection. Social SciSoc Sci Comput Rev 2005;23(4):486-501. [doi: 10.1177/0894439305278972]
41. Cunningham JA, Khadjesari Z, Bewick BM, Riper H. Internet-based interventions for problem drinkers: From efficacy trials to implementation. Drug Alcohol Rev 2010 Nov;29(6):617-622. [doi: 10.1111/j.1465-3362.2010.00201.x] [Medline: 20973846]
42. 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 Jun;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
43. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
44. Petry NM, Martin B, Cooney JL, Kranzler HR. Give them prizes, and they will come: contingency management for treatment of alcohol dependence. J Consult Clin Psychol 2000 Apr;68(2):250-257. [Medline: 10780125]
45. Postel MG, de Haan HA, ter Huurne ED, Becker ES, de Jong CAJ. Characteristics of problem drinkers in e-therapy versus face-to-face treatment. Am J Drug Alcohol Abuse 2011 Nov;37(6):537-542. [doi: 10.3109/00952990.2011.600388] [Medline: 21797813]

Abbreviations

CI: confidence interval
DASS-21: Depression Anxiety Stress Scale
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th revision
GHQ-28: 28-item General Health Questionnaire

XSL•FO
RenderX

MAP-HSS: Maudsley Addiction Profile-Health Symptom Scale
MfT: TCU Motivation for Treatment scale
RCQ: Readiness to Change Questionnaire
RCT: randomized controlled trial
ROC: receiver operating characteristic

Edited by G Eysenbach; submitted 18.03.11; peer-reviewed by H Riper, S Rooke; comments to author 09.05.11; revised version received 04.07.11; accepted 25.09.11; published 27.12.11.

Please cite as:
Postel MG, de Haan HA, ter Huurne ED, van der Palen J, Becker ES, de Jong CAJ
Attrition in Web-Based Treatment for Problem Drinkers
J Med Internet Res 2011;13(4):e117
URL: http://www.jmir.org/2011/4/e117/
doi:10.2196/jmir.1811
PMID:22201703

©Marloes G Postel, Hein A de Haan, Elke D ter Huurne, Job van der Palen, Eni S Becker, Cor AJ de Jong. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Virtual Reality to Provide Health Care Information to People With Intellectual Disabilities: Acceptability, Usability, and Potential Utility

Valerie Hall[1], MSc, DPhil; Suzanne Conboy-Hill[2], BA(Hons), MSc, MPhil, PhD; Dave Taylor[3], BSc, MSc

[1]Centre for Health Research, University of Brighton, Brighton, United Kingdom

[2]Research and Development, Sussex Partnership NHS Foundation Trust, Hove, United Kingdom

[3]Division of Surgery, Imperial College, London, United Kingdom

Corresponding Author:
Valerie Hall, MSc, DPhil
Centre for Health Research
University of Brighton
265 Mayfield House
Falmer
Brighton, BN1 9PH
United Kingdom
Phone: 44 1273 644015
Fax: 44 1273 644541
Email: v.hall@brighton.ac.uk

Abstract

Background: People with intellectual disabilities have poor access to health care, which may be further compromised by a lack of accessible health information. To be effective, health information must be easily understood and remembered. People with intellectual disabilities learn better from multimodal information sources, and virtual reality offers a 3-dimensional (3D) computer-generated environment that can be used for providing information and learning. To date, research into virtual reality experiences for people with intellectual disabilities has been limited to skill-based training and leisure opportunities within the young to mid age ranges.

Objective: This study assessed the acceptability, usability, and potential utility of a virtual reality experience as a means of providing health care-related information to people with intellectual disabilities. We designed a prototype multimodal experience based on a hospital scenario and situated on an island in the Second Life 3D virtual world. We wanted to know how people of different ages and with varying levels of cognitive function would participate in the customized virtual environment, what they understood from being there, and what they remembered a week later.

Methods: The study drew on qualitative data. We used a participatory research approach that involved working alongside people with intellectual disabilities and their supporters in a community setting. Cognitive function was assessed, using the Matrix Analogies Test and the British Picture Vocabulary Scale, to describe the sample. Participants, supported by facilitators, were video recorded accessing and engaging with the virtual environment. We assessed recall 1 week later, using a specialized interview technique. Data were downloaded into NVivo 8 and analyzed using the framework analysis technique.

Results: Study participants were 20 people aged between 20 and 80 years with mild to severe intellectual disabilities. All participants were able to access the environment and voluntarily stayed there for between 23 and 57 minutes. With facilitator support, all participants moved the avatar themselves. Participants engaged with the scenario as if they were actually there, indicating cognitive presence. Some referred back to previous medical experiences, indicating the potential for experiential knowledge to become the foundation of new learning and retention of knowledge. When interviewed, all participants remembered some aspects of the environment.

Conclusions: A sample of adults with intellectual disabilities of all ages, and with varying levels of cognitive function, accessed and enjoyed a virtual-world environment that drew on a health care-related scenario, and remembered aspects of it a week later. The small sample size limits generalizability of findings, but the potential shown for experiential learning to aid retention of knowledge on which consent is based appears promising. Successfully delivering health care-related information in a non-National

XSL•FO
RenderX

Health Service setting indicates potential for delivery in institutional, community, or home settings, thereby widening access to the information.

(J Med Internet Res 2011;13(4):e91) doi:10.2196/jmir.1917

KEYWORDS

Learning disabilities; intellectual disabilities; virtual reality; health information; participatory research; capacity to consent; presence

Introduction

People with intellectual disabilities have the poorest access to health care [1], which may be worsened by a lack of accessible health information. In the United Kingdom, the Mental Capacity Act [2] makes it a legal requirement for health professionals to ensure patients are given full information, to enable them to make their own decision about treatment. However, people with intellectual disabilities may have difficulties in taking in and retaining information in order to make that decision, and therefore they may not get the treatment they need, or they may get treatment they didn't want. Even when they agree to treatment, if they do not fully understand what is going to happen to them, they may refuse to cooperate. This could be distressing for the person, his or her caregiver, and the health care staff, and may lead to a longer stay in hospital.

Normally, information leaflets and storybooks are used to provide health care information. However, a review of informed consent to health care interventions concludes that enhancing understanding may depend on the effort made to tailor the information to the abilities and needs of the individual with intellectual disabilities [3]. It is already known that learning in people with intellectual disabilities can be enhanced by using audio and video presentations [4,5]. Moreover, interactive multimedia technologies such as virtual reality provide opportunities for people to interact with virtual objects and events from everyday life, which can lead participants to feel that they are actually "there"—a subjective experience known as cognitive presence [6]. These techniques may help bridge the gap between information representation and experiential learning [7]. Virtual reality has been shown to support learning in people with intellectual disabilities in a variety of ways [8-11] and to provide a safe setting in which they can practice activities that might not be possible in the real world [12].

Gaming technology, which can enhance motivation, is being used increasingly to develop interventions that improve health knowledge and assist in health-related decision making for the general population [13]. Virtual reality gaming studies show that people with intellectual disabilities enjoy experiences that allow them to take control of their environment and succeed in activities that are usually inaccessible to them [14]. Use of Internet-based virtual reality environments, such as Second Life, is also increasing. These environments can be accessed from any location and offer unique and interactive ways to facilitate health care information, particularly when full advantage is taken of the experiential features [15].

Existing research conducted with people with intellectual disabilities, using virtual reality applications, mainly relates to skill-based training [9,10,16-18], rehabilitative skills [12,19], developing participation in exercise skills [20], or leisure activity [14], and is mostly undertaken with a younger group of people in institutional settings. Our study adds to existing knowledge because it reports on the acceptability, usability, and potential utility of virtual reality as a means of providing health care-related information to people with intellectual disabilities, and includes adult participants from the whole age range, including older people. Delivering health care-related information in a social setting indicates potential for its use in community or home settings, thereby widening peoples' access to it. Importantly, we used a participatory research method, working alongside people with intellectual disabilities and their supporters to ensure their rights were recognized within the process and their experiences were properly represented [21].

Methods

Study Design

This exploratory study drew on qualitative data, to assess the acceptability, usability, and potential utility of a virtual reality experience to provide health care-related information to people with intellectual disabilities. We were interested in how people of different ages and with varying levels of cognitive function would participate in the customized virtual environment, what they understood from being there, and what they remembered a week later.

We wanted to make the research participatory by working alongside people with intellectual disabilities and their supporters. Therefore, we worked collaboratively with the Grace Eyre Foundation in Hove, East Sussex, UK, which is a registered charity, providing support and services in the community for people with intellectual disabilities. We also involved people with intellectual disabilities in the delivery of the study itself, taking care to remain mindful of the potential vulnerability of the participants, and the need to pay close attention to issues of recruitment and consent. The conduct of the research was overseen by a steering group, which included a person with intellectual disabilities and representatives from Grace Eyre. Not only did this provide a reference group within which to discuss and monitor ethical practice, but also members provided practical advice on the production of recruitment information.

The participants were video recorded while accessing and engaging with the virtual environment. Six volunteer psychology graduates, who had been given rudimentary instruction on the use of Second Life (http://secondlife.com/) and the navigational tools, acted as facilitators to the delivery of the virtual environment. They encouraged the participants to recognize

XSL•FO
RenderX

landmarks, and to engage and experiment with various interactive elements contained within the environment.

We assessed each participant's memory of the virtual environment exposure 1 week later using a modified cognitive interview technique. The cognitive interview is made up of two parts. In the first part, the participant is asked to recount as much as possible of the experience, without interruption. The interviewer then probes the information systematically, using contemporaneous notes. Nonleading questions target key items of information; other questions probe meaning. In the second part, participants are shown screenshots of the exposure and asked questions aimed at further prompting memory. This technique has been shown to increase the reporting of accurate recall from various population groups [22-24], particularly if, as in this study, staged events are used [25]. An experienced clinical psychologist conducted the modified cognitive interview, with a 1-week time lapse between exposure and interview to mirror usual clinical practice when assessing retention of information on which capacity to consent is assessed.

We chose to video record the participants accessing and engaging with the virtual environment, although we acknowledge the potential criticisms of observer bias in studies using observational data. However, the communication difficulties associated with our participant group, and their recognized tendency to agreement and compliance, precluded direct questioning through standard interview or questionnaire. Data triangulation, achieved through comparing the results of the observation analysis with cognitive interview data and a focus group validation of findings, served to support study credibility [26].

Setting and Participants

The research was conducted at the Grace Eyre social center, which provided a well-resourced information technology suite. Undertaking the research at the center, rather than in a laboratory, allowed us to assess the potential for delivering the virtual environment in a real-life community setting. Moreover, it provided a sense of familiarity and security to the research participants. Participants were invited to bring their support workers, but only 7 chose to do so.

We recruited a convenience sample from people with intellectual disabilities who use the center. The sample comprised 20 people, 11 male and 9 female, between the ages of 20 and 80 years.

In order to describe the level of cognitive function in the sample, we used two minimally demanding screening tools: one for verbal material, and the other visual. The British Picture Vocabulary Scale (BPVS) assesses contextual receptive vocabulary and does not rely on speech or reading. The BPVS test is acceptable to adults with intellectual disabilities because there is minimal experience of failure. The Matrix Analogies Test (MAT) is a similar tool, requiring little language and no writing skills. Scoring for each is governed by clear, manualized criteria, and gives rise to raw, age-equivalent, and standard scores, from which an intelligence quotient can be estimated. From the analysis of the test results we determined that the sample consisted of people with low, medium, and high cognitive function within the intellectual disabilities range, with one exception indicated in the MAT score, a person with an autistic spectrum disorder (see Table 1). In addition, we observed other conditions such as Down syndrome.

Of the study participants, 16 had previously used a computer at Grace Eyre, but only 7 had medium and 1 had high levels of computer usage. Although the computers in the information technology suite were set up for use by people with intellectual disabilities, they had not previously been used to access Second Life. People with intellectual disabilities may have trouble operating multifunction control devices due to problems in remembering which device achieves which task, or they may experience fine-motor difficulties, which could leave them feeling frustrated and demotivated [27]. Using the equipment in the center, we undertook a small prestudy test to identify and rectify any preliminary problems with the navigational and interaction control devices used to access the virtual environment. No adjustments to the equipment were required.

Table 1. Levels of cognitive function within the sample of people with intellectual disabilities

BPVS[a] age equivalent in years	Number of people in range (N = 20)	MAT[b] standard score	Number of people in range (N = 20)
Low (0–4)	4	Low (≤49)	7
Medium (5–6)	14	Medium (50–75)	12
Highest (10–14)	2	Highest (76–100)	1

[a] British Picture Vocabulary Scale.

[b] Matrix Analogies Test.

Consent

We paid particular attention to obtaining consent in this potentially vulnerable population. People with intellectual disabilities helped us write the patient information sheet and consent forms (Multimedia Appendix 1, Multimedia Appendix 2). Posters were displayed in the Grace Eyre center with an invitation to contact a staff member for more information and support to decide whether to volunteer (Multimedia Appendix 3). Center staff sought initial consent from the participant but verbal consent was also sought immediately prior to each element of the study.

Approval to undertake the study was granted by the University of Brighton Research Ethics and Governance Committee. In addition, UK National Health Service Research Governance approval was granted.

XSL•FO
RenderX

Virtual Environment Health Information Experience

A virtual environment, representing a stylized hospital building and internal rooms, was designed by Imperial College London. The environment was hosted on a private Second Life server and was accessible to the study, but not to the public, on a desktop computer over the Internet. Access to the virtual environment was limited to authorized account holders. Authorized users could login and tour the environment while being represented as an animated avatar in male or female human form. Users were able to control their avatars' movements using a computer keyboard's arrow keys and a mouse. A realistic 3-dimensional rendering of the key landmarks around the hospital in Brighton was developed, and suitable streets that could be easily navigated were linked to these (see figure 1. A realistically animated ocean bounded the simulated environment, and distant views of cliffs and buildings were produced using large photographic images placed on the borders of the simulated space, rather like a film set. Two avatars were created with features that could be selected to provide a broad match with the participant, such as male or female, hair color, and ethnicity, plus a wheelchair if required.

Interiors of the hospital buildings were also replicated (see figure 2). The hospital scenario incorporated a programmed "nurse" robot, which was activated by participants taking a seat in a waiting room. The nurse communicated with the participants using preprogrammed text in a dialogue box and could be summoned by phone to a specific room in order to explain its purpose. In addition, the nurse offered a tour of the hospital, with the participants having the choice of being pushed on a hospital bed, using a wheelchair, or walking from room to room. The nurse robot was generated by a special version of the Second Life client software running on a virtual server in the Amazon "cloud" and controlled by a script that specified its responses and actions.

The hospital also included a clinical examination room, with a bed on which the participant was invited to lie, and could experience an interactive blood pressure machine and cuff (see figure 3, a preparation room, an operating room wherein the participant could lie on the operating table, and a recovery ward containing a static patient in a bed.

Other interactive hospital equipment was built into the rooms to provide a suitably realistic experience. All of these virtual-world objects were automated using programs written in the Second Life java-like scripting language (Linden Scripting Language). Multimedia Appendix 4 provides a video walk-through of the virtual environment.

Figure 1. Screenshot of opening scene showing key landmarks and standard female avatar.



XSL•FO
RenderX

Figure 2. Screenshot of hospital waiting room.



Figure 3. Screenshot of clinical examination room.



Data Capture, Production, and Analysis

The participants were video recorded while accessing and engaging with the virtual environment. A portable usability lab designed by the University of Brighton enabled us to capture audiovisual data in the community center. Two high-definition camcorders were strategically positioned: one to capture information about the physical use of computer equipment and navigational tools, and the other to capture the participants' engagement with, and response to, the scenario on the computer

XSL•FO
RenderX

screen. A video scaler and recorder captured concurrent images from the computer screen. Data from all three streams were merged using Apple iMovie 09 (Apple, Cupertino, CA, USA) video-editing software and downloaded into NVivo version 8 (QRS International, Doncaster, Australia), where an annotated account of key events in each exposure video was prepared.

Cognitive interviews were digitally recorded and transcribed verbatim. To determine the degree of accurate recall, a summary of each individual's virtual environment experience was compared with the events they recalled in the cognitive interview. This was undertaken by the researcher who analyzed the video exposures, to reduce the risk of interviewer bias.

All data were entered into NVivo and analyzed the using the framework approach, which allows for both deductive and inductive analysis [28]. This approach involves a systematic process of sifting, charting, and sorting material according to key issues and themes, which is appropriately targeted toward providing "answers" in the form of greater illumination or understanding of the issues. As there were specific questions that we wanted to answer, the framework for analysis was formed from themes arising directly from these questions and from the literature (see Table 2).

Textual data were analyzed thematically, with key issues arising from the analysis being entered into a casebook matrix in order to compare results across different characteristics or within and across themes (Multimedia Appendix 5). For instance, from the cross-analysis we were able to report on whether the people who enjoyed the experience were more likely to have identified with the avatar.

Table 2. Major themes and issues underpinning the framework analysis

Major themes	Related issues
Temporality	Time taken to sit at computer, use keyboard, gain ability to move avatar, recognize scenario, identify with the avatar; overall time spent in Second Life and looking at the screen
Accessibility	Technical and physical barriers; chair and body position in relation to screen; moving and interacting in the virtual space; facilitators' support of skills development
Context	Physical characteristics of user compared with avatar; recognition of scenario; facilitation style; role of support worker; culture of Grace Eyre; users' previous experience of computers
Cognitive presence	Reaction to user interface, authentic scenario; identification with scenario; structure of scenario; visual cues; engagement with interactive activities; spontaneous recall of previous experience; relationship to avatar; emotional response to avatar actions; engagement with "nurse" robot; engagement with chat-based interaction; control; autonomy; enjoyment
Recall	Number of accurate statements, distortions, confabulations, and imported information

Results

We set out to explore the acceptability, usability, and potential utility of virtual reality as a means of providing health care-related information to people with intellectual disabilities. All 20 participants completed the Second Life exposure and were interviewed, using the cognitive interview technique, within a 5-month period. The results from both these data sources are presented using the major theme headings outlined in Table 2.

Temporality

All participants sat at the computer, began to engage with the virtual environment immediately, and maintained good concentration throughout, only disengaging for short tea or toilet breaks. Within 5 minutes of starting the virtual environment exposure, 18 people independently moved the avatar, with the other 2 taking up to 10 minutes to do so. A total of 17 participants recognized aspects of the Brighton scenario right away. All participants instantly noticed the avatar, and almost all of them thought that it was "pretending" to be them. They remained in the exposure voluntarily between 23 and 57 minutes, with the majority staying between 40 and 45 minutes.

Accessibility

None of the participants had physical difficulty using the keyboard, although skill levels varied. Most of those who started with a low level of keyboard skills improved during the exposure. However, there were a small number whose skills remained underdeveloped throughout. With facilitator support, a lack of skills did not seem to inhibit either active engagement in the scenario or decisions as to where the avatar went and what it did.

The following extract from an annotated video record illustrates this observation:

> *[Participant] understands how to move the avatar but reluctant to do so on her own without support although later in this room her fingers hover over keys in readiness to press them before guided. Gets on bed with help. Stands on bed by mistake and finds this funny.*

The facilitators acted mainly as guides but sometimes intervened when participants could have done things for themselves. The ability to move the avatar skillfully and autonomously meant that sometimes people concentrated more on movement than on the content of the virtual environment, as demonstrated in the following annotated video record:

> *Doesn't know the purpose of the waiting room. Seeks help to move avatar onto seat. Not interested in the dialogue with the nurse or even acknowledging her presence other than to move the avatar to the next room.*

XSL•FO
RenderX

Context

Two avatars were created with features that could be chosen to provide a broad match with the participant, such as male or female, hair color, and ethnicity, plus a wheelchair if required. However, other than gender, and skin color to reflect ethnicity, none of the other features were used in the study.

As a result, only 4 participants looked similar to their avatar and therefore it is difficult to assess whether this affected identification with it.

Everyone recognized some aspects of the virtual environment, although not everyone identified the outside of the hospital or was aware of the purpose of all the rooms they visited. The operating room was the most likely to be recognized spontaneously during the virtual environment experience, followed by the waiting room. However, in the cognitive interview, participants remembered activities that occurred in the assessment room, such as the use of the blood pressure machine.

The facilitators were instructed to enable a nonthreatening experience using self-directed, informal, and playful strategies. A lighthearted approach is important because engaging emotionally with the virtual environment improves enjoyment, thereby aiding memory. However, observed styles ranged from enabling to directive across the participant group, and sometimes even within the individual's virtual environment experience. An enabling style encouraged participants to go where they wanted and do what they wanted, but it did mean that the balance of the exposure was lost, and the person often did not spend time in every area. Additionally, it meant that they might not have had the opportunity to experience many of the activities. A more directive style often meant the participant spent time in all the areas, but it sometimes resulted in the facilitator taking over the controls or telling the participant where to visit, which lost some of the opportunities for playful engagement.

The ideal facilitation style appears to be one that enables access to all the opportunities available in the virtual environment but takes a negotiated approach to determining what the person will do in it, and how much help they need and want. However, it should also be gently directive in supporting the person to learn to move the avatar to the best of his or her ability, while spending sufficient time in each area. We term this assertive facilitation.

Support worker involvement was minimal, occasionally sharing jokes and enjoyment and offering encouragement. Only 1 participant needed continuous input from her support worker to help her stay focused.

Our participants were recruited from an organization that takes a positive stance toward its clients' use of technology and provides a structured day that encourages focused activity. This positive environment and previous experience of using computers may have influenced the length of time participants voluntarily stayed at the keyboard and their willingness to engage in the experience.

Cognitive Presence

Participants clearly knew that they were interacting with the virtual environment through a computer because they were using the keyboard and mouse to access it, but this did not detract from their engagement with the scenario. Several conditions promoted a sense of cognitive presence.

All participants identified with one or more of the virtual environment areas, with 17 recognizing Brighton sea front instantly. This led to a high degree of engagement, as the participants recalled previous experiences while in the virtual environment, expecting to see boats and to go swimming, and even expecting to be able to locate their own home. Most people were curious and explored the outdoor, as well as the indoor, environment. They tried to open doors in buildings to find out what was behind them; they rang doorbells and tried to sit on seats. Within the hospital component of the virtual environment they happily engaged in the programmed activities, such as having their blood pressure taken, or lying on the operating room table. One participant said "And I laid on a bed to see what it feels like when you do have an operation."

More importantly, the experience also prompted them to recount prior experiences of hospital treatments, such as having blood or blood pressure taken, or x-rays, or more generally about being in hospital. When recognizing an x-ray machine, one participant said "I saw them before...when I had my hip done."

These associations were so strong that some people recounted them in the interview a week later. The ability to extrapolate information from the virtual environment is important, as it indicates understanding of the scenario and provides a foundation for learning about it [29].

Full identification with the avatar did not seem to be important; only a few participants referred to the avatar as "I." Others had a more superficial relationship with it, using it as a way of navigating through the scenario, with some expressing concern for what it was doing in potentially dangerous situations—for example, when crossing the road.

The technical performance of the nurse robot proved unpredictable during the virtual environment experience: some participants followed her and some verbalized recognizing her, but only one person wanted to talk to her. However, in the interview a week later, a large number of people mentioned a nurse and the patient in the bed.

> *Took him (the avatar)...um...yes...nurse going there—and I go—and (the nurse—dialogue spoken by the facilitator) asked me about it—about being in the hospital there.*
>
> *There was some writing on the screen...um—I am better now and I will go home...yes—that is what the patient said.*

This indicates the potential value of including a human-like "other" presence with a specific purpose appropriate to the scenario.

A total of 18 participants moved the avatar themselves or determined what it did. If the facilitator intervened too much, or control was taken away, some participants appeared to lose

XSL•FO
RenderX

interest or confidence. In one example, a participant who chose to have his avatar pushed around the hospital on the bed epitomizes this loss of confidence. Prior to the bed tour the participant was attempting to move the avatar himself, but following the bed tour, he required continual prompting before he returned to his previous activity level. In another example, the facilitator takes over control of the keys, and the participant is observed losing concentration and starting to look round the information technology suite.

Maintaining control of the avatar allowed participants to satisfy their curiosity by going to look at things spontaneously. One participant asked "What's that in there?" and, encouraged by the facilitator's prompt "Let's go and have a look—get a bit closer", engaged in a dialogue about the use of the piece of hospital equipment.

The ability to move the avatar skillfully and autonomously was also linked with enjoyment. However, even those who had not improved their skills expressed enjoyment during the experience and also in the cognitive interview afterward, sometimes asking to repeat the virtual environment experience. For example, in response to the cognitive interview question "What else happened in the computer?" one participant with undeveloped skills responded "Dun a 'puter myself" and later in the interview the participant reiterated "Do it again," and yet later "Do it again—what date?"

Even when skills were underdeveloped, participants still wanted to navigate the avatar themselves and expressed enjoyment at the result, prompting exclamations such as:

> *Because you can look up the hospital.You can press what you can do on those little—thing—pointer things. On the computer. You can. Yes I enjoyed looking on the computer.*

Judged through observations of their facial expressions and body language during the exposure, 17 participants demonstrated enjoyment—for instance, smiling, leaning forward into the screen, laughing and pointing to elements on screen, and commenting and making jokes. Although we cannot say whether the other 3 participants enjoyed the experience because their expressions remained neutral throughout, they did not show any signs of physical or emotional agitation such as rocking, stereotypy, or distractibility. In fact, all 3 stayed in the virtual environment voluntarily for over 30 minutes.

From the above it can be seen that enjoyment was linked to the recognition of the scenario, the sense of achievement in moving or directing the avatar, or engagement in the activities, which sometimes stimulated wonder and amusement. When asked to explain, in the cognitive interview, why she had said the experience was quite good, one participant said:

> *Well, going into hospital and look around. They give you confidence and then you won't be frightened when you do go in.*

Recall

Participants were interviewed 1 week after the virtual environment experience. The cognitive interview is made up of two parts. In the first part, the participant was asked to recount as much as possible of the experience, without interruption. Nonleading questions targeted key items of information and asked—for instance, "You said you stood in the operating theatre, tell me what you saw in the operating theatre." Other questions probed meaning, as in "You said there was an operating table; tell me what that's for, what happens there." In the second part of the interview, participants were shown screenshots of the exposure and were asked questions aimed at further prompting memory (for examples of screenshots see Figure 1, Figure 2, and Figure 3).

All participants reported some accurate memories, but the amount varied. Both parts of the interview elicited some made-up (confabulatory) information, or information that was added from their own experiences. This was significantly higher when memory was prompted by screenshots, with some participants "remembering" information about parts of the scenario that they had not visited. Although the inclusion of confabulated information is not unusual in the nonintellectual disabilities population, the increased reporting in the screenshot section indicates that the interview procedure needs to be revised. However, although our primary aim was to test for accuracy, it is important to bear in mind that the confabulated information was often based on the participant's own personal experiences, indicating that they had an understanding of what was in the virtual environment and demonstrating potential for building on this knowledge.

Validation of the Findings

At the end of the study, 2 months after the cognitive interview, 8 participants volunteered to take part in a 35-minute structured focus group to validate the results. The clinical psychologist and another member of the research team with experience of working with marginalized groups led the participants through the findings. To enhance the credibility of this process, we invited one of the steering group members, a person with intellectual disabilities, to observe the focus group.

The participants confirmed that the results matched their perceptions of the virtual environment experience. They all offered unprompted comments about various aspects of the exposure, and many were keen to be involved in further stages of the virtual environment development, with some suggesting other things that could be added to the scenario. At the end of the focus group, our steering group member spontaneously commented that he was impressed that the participants were still "buzzing" about the experience 2 months later.

Discussion

The study found that adults with intellectual disabilities of all ages, and with various levels of cognitive function, could access and enjoy a virtual environment that drew on a health care-related scenario. They engaged with the scenario as if they were actually there, which encouraged them to talk about previous experiences. This indicates a potential for experiential knowledge to become the foundation of new learning about a health care-related scenario. All participants were able to remember some aspects of the virtual environment when interviewed a week later.

XSL•FO
RenderX

We were surprised by the degree of concentration and engagement shown by the participants. One of the influencing factors might have been the high degree of scenario recognition, through which the participants appeared to exhibit the subjective sensation of feeling and behaving as if they were actually there [30]. This engagement with the scenario is akin to the concept of cognitive presence, the predeterminants of which have been attributed variously to facilitation by technological equipment [7,31], or its psychological [32] or multidimensional nature [6]. Heeter's [6] 3-dimensional model, consisting of personal presence (the sense of being there), social presence (reaction to other beings in the virtual environment), and environmental presence (the extent to which the environment appears to know that the person is there), best reflects the way we designed the environment. Our observations of the participants' responses to the environment, in terms of their relationship to the nurse and patient, plus the interactive activities, support the relevance of the components of Heeter's model. However, our participants were also aware of the need to use navigational tools to access the experience, but this did not seem to undermine their sense of being there, contrary to other authors' suggestions that the sense of presence is greater when the mediation process remains unnoticed by the users [33]. Heeter [34] argues that other factors affect presence in nonmediated situations, and further consideration of these might help conceptualization of mediated presence. Thus, our findings support Thornson and colleague's [32] arguments for presence as a phenomenon occurring in the mind, rather than in the specific technology. Human factors may also determine a person's tendency to experience the cognitive state of presence [35]. Although the factors, described by Thornson et al [32] as empathy, spatial orientation, cognitive involvement (both passive and active), ability to construct mental models, and introversion, were identified using questionnaires with college graduates, they demonstrate some resonance with the observations made of the participants in this study. However achieved, the subjective experience of being there prompted participants in this study to remember previous experiences of medical treatments and visits to a hospital, which has the potential to provide the foundation for new learning, to open up an opportunity for clinical dialogue in order to elicit additional clinical information, and to assess psychosocial concerns [36].

These findings have implications for the mode of information delivery required to enhance the assessment of capacity to consent in people with intellectual disabilities and other marginalized groups. Del Carmen and Joffe [37] indicate five elements necessary for valid informed consent: voluntarism, capacity, disclosure, understanding, and decision. Assessment of capacity depends on the quality of the information provided, and this study has examined the feasibility of using a virtual environment as a way of *disclosing* information to people with intellectual disabilities in a way that enables them to *understand* the information and its relevance to their own situation. It is clear that the people in this study could access the virtual environment, engage with it for long enough to understand what it represented, and remember information about it a week later, mirroring the time lapse between giving information and interviewing to assess capacity that occurs in actual practice.

Much of the research regarding consent in vulnerable populations relates to ability to recall information [38,39] or to make decisions [40]; however, there are also issues of ongoing consent, which have yet to be addressed [41]. Using a virtual environment to provide information to enable valid consent means it could be accessed and used freely, not only as a way of providing information on which the individual is assessed to have capacity to consent, but also, after initial consent, to ensure ongoing consent. Similarly, the opportunity to practice being a patient before coming into hospital may provide an increased sense of control over health care experiences [15].

In this study, psychology graduates facilitated access to the health care information and, although they had limited expertise in working with people with intellectual disabilities and no previous knowledge of Second Life, they needed little training to help participants access and navigate in Second Life. While we have commented on differing facilitation styles and speculated on how they might have influenced the participants' experience, this is largely because the virtual environment prototype was exploratory, related to a nonspecific health information event, and included greater opportunities for divergence from the health information purpose. A virtual environment designed to deliver health care information on a specific treatment would be more tightly structured, and therefore the balance between enabling and directive facilitation would change, depending on the purpose of its use and the role of the person providing the facilitation. This study indicates that the virtual environment could be delivered not only in health institutions and community environments, but also by caregivers and support workers as well as health care professionals. For instance, a physician could use the virtual environment initially to explain a potential treatment scenario, and this may require more directive facilitation, whereas a nurse or support worker may use the virtual environment to help understanding and reinforce the information. Additionally they might use it to help the person rehearse what might happen in a treatment situation, using a less directive approach, enabling the person to talk about previous experiences and possible concerns [42]. However, further work is required to ascertain the degree of skill required to avoid overcontrol and disengagement, or low control and inadequate exposure and therefore to maximize engagement. Our proposed next phase of the research intends to address this.

Limitations

Our participants were recruited from an organization that takes a positive stance toward its clients' use of technology and provides a structured day that encourages focused activity. Our small sample was recruited exclusively from this proactive organization, which precludes us from generalizing the findings to a wider population of people with intellectual disabilities. Moreover, we recognize that the scenario developed for this study was specifically tailored to our participants, which may have had a positive impact on the findings. While it may not be possible to customize future health care information experiences quite so specifically, new and emerging technological techniques and platforms, such as consumer-oriented 3-dimensional modeling, augmented reality, and high-fidelity 3-dimensional scanning, could enable some degree of "recognition."

XSL•FO
RenderX

We argued earlier that potential criticisms of interpreter bias regarding observation data have been offset by triangulating these data against the verbal responses made in the cognitive interview and by validating the findings in the focus group. However, we acknowledge that there may still have been some subjective bias, particularly when commenting on visual or behavioral expressions of enjoyment. In future work we would consider developing a specialized interview technique for gaining verbal and nonverbal accounts of participants' views.

Conclusion

Our study clearly demonstrates the potential for using virtual reality to provide health care-related information to people with intellectual disabilities. People with intellectual disabilities engaged with a health-related virtual environment experience as if they were actually there, which prompted them to talk about previous health care experiences. It also provided an opportunity for them to practice being patients, potentially providing more information about themselves and their worries, which could lead to an increase in confidence in treatment situations. Although we have not yet tested the effectiveness of the virtual environment model against existing 2-dimensional health care information delivery methods, the potential for experiential learning indicated in our study appears promising. Furthermore, successfully delivering health care-related information in social settings indicates potential for use in a variety of settings. Moreover, the study indicates the potential for several health-related applications such as use by physicians to explain treatments, or by nurses and support workers to help understanding and enable the person to rehearse what might happen in a treatment situation. Importantly, the opportunity to revisit the information-giving scenario offered by virtual environments may provide a way of addressing issues of ongoing consent.

Our study is the first step on a path to providing effective health information to people with intellectual disabilities, and we have learned a great deal by taking it. Our next step is to further develop the prototype with help from volunteers from our participant group. We will then test it out in a larger and more diverse population of people with intellectual disabilities and in a range of settings, drawing on the lessons learned in this exploratory study.

Acknowledgments

The article presents independent research commissioned by the National Institute for Health Research (NIHR) under the i4i programme award number II-FS-0908-10011. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

We would like to acknowledge the significant contribution of colleagues at the Grace Eyre Foundation, in particular Eva Jarvis and David Mathews. Thanks to Robin Winter at Imperial College London who designed the VE and also to Dr Richard Griffiths from the School of Computing, Engineering and Mathematics, University of Brighton who designed the mobile usability lab and to Heloise Candello who operated it. We would also like to thank the psychology graduates who facilitated the participants' experience. Finally thanks to Prof Angie Hart and Dr Christina Panton who commented on the development of this article.

Second Life is a trademark of Linden Research, Inc.

Conflicts of Interest

None declared

Multimedia Appendix 1

Patient information sheet.

[PDF File (Adobe PDF File), 67KB - jmir_v13i4e91_app1.pdf]

Multimedia Appendix 2

Consent form.

[PDF File (Adobe PDF File), 22KB - jmir_v13i4e91_app2.pdf]

Multimedia Appendix 3

Recruitment poster.

[PDF File (Adobe PDF File), 408KB - jmir_v13i4e91_app3.pdf]

Multimedia Appendix 4

Video walk-through of the virtual environment.

[MOV File, 44MB - jmir_v13i4e91_app4.mov]

XSL•FO
RenderX

Multimedia Appendix 5

NVivo 8 casebook analysis matrix showing attributes.

[XLS File (Microsoft Excel File), 28KB - jmir_v13i4e91_app5.xls]

References

1. Mencap. 2007. Death by Indifference: Following up the Treat Me Right! Report URL: http://www.mencap.org.uk/sites/default/files/documents/2008-03/DBIreport.pdf [accessed 2011-08-18] [WebCite Cache ID 6122AGrQS]
2. UK HM Government. UK National Archives. 2007 Jul 19. Mental Health Act 2007, Chapter 12 URL: http://www.legislation.gov.uk/ukpga/2007/12/pdfs/ukpga_20070012_en.pdf [accessed 2011-08-22] [WebCite Cache ID 618A4hy9P]
3. Goldsmith L, Skirton H, Webb C. Informed consent to healthcare interventions in people with learning disabilities--an integrative review. J Adv Nurs 2008 Dec;64(6):549-563. [doi: 10.1111/j.1365-2648.2008.04829.x] [Medline: 19120568]
4. Dunn A, Kroese BS, Thomas G, Mcgarry A, Drew P. “Are you allowed to say that?” Using video materials to provide accessible information about psychology services. Br J Learn Disabil 2006;34(4):215-219. [doi: 10.1111/j.1468-3156.2005.00373.x]
5. Broughton S. A review of the literature: interventions to maximize capacity to consent and reduce anxiety of women with learning disabilities preparing for a cervical smear test. Health Serv Manage Res 2002 Aug;15(3):173-185. [Medline: 12184837]
6. Heeter C. Being there: the subjective experience of presence. Presence (Camb) 1992;1:262-271 [FREE Full text]
7. Winn W. Human Interface Technology Laboratory, Washington Technology Center, University of Washington. 1993 Aug. A Conceptual Basis for Educational Applications of Virtual Reality URL: http://www.hitl.washington.edu/publications/r-93-9/ [accessed 2011-08-22] [WebCite Cache ID 618AJORm8]
8. Parsons S, Mitchell P. The potential of virtual reality in social skills training for people with autistic spectrum disorders. J Intellect Disabil Res 2002 Jun;46(Pt 5):430-443. [Medline: 12031025]
9. Cromby JJ, Standen PJ, Newman J, Tasker H. Successful transfer to the real world of skills practised in a virtual environment by students with severe learning difficulties. In: Proceedings. 1996 Presented at: First European Conference on Disability, Virtual Reality and Associated Technologies; July 8-10, 1996; Maidenhead, UK URL: http://playpen.icomtek.csir.co.za/~acdc/assistive%20devices/Artabilitation2008/archive/1996/papers/1996_12.pdf
10. Cromby JJ, Standen PJ, Brown DJ. The potentials of virtual environments in the education and training of people with learning disabilities. J Intellect Disabil Res 1996 Dec;40 (Pt 6):489-501. [Medline: 9004109]
11. Alm N, Arnott JL, Murray IR, Buchanan I. Virtual reality for putting people with disabilities in control. 1998 Presented at: IEEE International Conference on Systems, Man, and Cybernetics; October 11-14, 1998; Orlando, FL, USA. [doi: 10.1109/ICSMC.1998.727865]
12. Schultheis MT, Rizzo AA. The application of virtual reality technology in rehabilitation. Rehabil Psychol 2001;46(3):296-311. [doi: 10.1037/0090-5550.46.3.296]
13. Reichlin L, Mani N, McArthur K, Harris AM, Rajan N, Dacso CC. Assessing the acceptability and usability of an interactive serious game in aiding treatment decisions for patients with localized prostate cancer. J Med Internet Res 2011;13(1):e4 [FREE Full text] [doi: 10.2196/jmir.1519] [Medline: 21239374]
14. Yalon-Chamovitz S, Weiss PL. Virtual reality as a leisure activity for young adults with physical and intellectual disabilities. Res Dev Disabil 2008 Jun;29(3):273-287. [doi: 10.1016/j.ridd.2007.05.004] [Medline: 17590313]
15. Beard L, Wilson K, Morra D, Keelan J. A survey of health-related activities on second life. J Med Internet Res 2009;11(2):e17 [FREE Full text] [doi: 10.2196/jmir.1192] [Medline: 19632971]
16. Rose FD, Brooks BM, Attree EA. An exploratory investigation into the usability and usefulness of training people with learning disabilities in a virtual environment. Disabil Rehabil 2002;24(11-12):627-633. [doi: 10.1080/09638280110111405] [Medline: 12182803]
17. Brooks BM, Rose FD, Attree EA, Elliot-Square A. An evaluation of the efficacy of training people with learning disabilities in a virtual environment. Disabil Rehabil 2002;24(11-12):622-626. [doi: 10.1080/09638280110111397] [Medline: 12182802]
18. Groenewegen S, Heinz S, Frohlich B, Huckauf A. Virtual world interfaces for special needs education based on props on a board. Comput Graph 2008;32(5):589-596. [doi: 10.1016/j.cag.2008.07.002]
19. Standen PJ, Brown DJ. Virtual reality in the rehabilitation of people with intellectual disabilities: review. Cyberpsychol Behav 2005 Jun;8(3):272-82; discussion 283. [doi: 10.1089/cpb.2005.8.272] [Medline: 15971976]
20. Lotan M, Yalon-Chamovitz S, Weiss PL. Lessons learned towards a best practices model of virtual reality intervention for individuals with intellectual and developmental disability. New York: IEEE; 2009 Presented at: Virtual Rehabilitation International Conference, 2009; June 29-July 2, 2009; Haifa, Israel. [doi: 10.1109/ICVR.2009.5174208]
21. Walmsley J. Involving users with learning difficulties in health improvement: lessons from inclusive learning disability research. Nurs Inq 2004 Mar;11(1):54-64. [Medline: 14962347]
22. Memon A, Bull R. The cognitive interview: its origins, empirical support, evaluation and practical implications. J Commun Appl Soc Psychol 1991;1(4):291-307. [doi: 10.1002/casp.2450010405]

XSL•FO
RenderX

23. Milne R, Bull R. Back to basics: a componential analysis of the original cognitive interview mnemonics with three age groups. Appl Cogn Psychol 2002;16(7):743-753. [doi: 10.1002/acp.825]
24. Conboy-Hill S. Vulnerable adults: assessing capacity to consent. Clin Psych Forum 2006;158:19-23.
25. Köhnken G, Milne R, Memon A, Bull R. The cognitive interview: a meta-analysis. Psychol Crime Law 1999;5(1-2):3-27. [doi: 10.1080/10683169908414991]
26. Guba E, Lincoln Y. Fourth Generation Evaluation. Newbury Park, CA: Sage Publications; 1989.
27. Standen PJ, Brown DJ, Anderton N, Battersby S. Systematic evaluation of current control devices used by people with intellectual disabilities in non-immersive virtual environments. Cyberpsychol Behav 2006 Oct;9(5):608-613. [doi: 10.1089/cpb.2006.9.608] [Medline: 17034329]
28. Richie J, Spencer L. Qualitative data analysis for applied policy research. In: Huberman AM, Miles M, editors. The Qualitative Researcher's Companion. Thousand Oaks, CA: Sage Publications; 2002:305-329.
29. Knowles M. The Modern Practice of Adult Education: From Pedagogy to Andragogy. Wilton, CT: Association Press; 1980.
30. Barfield W, Zeltzer D, Sheridan T, Slater M. Presence and performance within virtual environments. In: Barfield W, Furness III T, editors. Virtual Environments and Advanced Interface Design. New York, NY: Oxford University Press; 1995:473-513.
31. Biocca F, Levy MR. Communication in the Age of Virtual Reality. Hillsdale, NJ: L Erlbaum Associates; 1995.
32. Thornson CA, Goldiez BF, Le H. Predicting presence: constructing the tendency toward presence Inventory. Int J Hum Comput Stud 2009 Jan;67(1):62-78. [doi: 10.1016/j.ijhcs.2008.08.006]
33. Lombard M, Ditton T. At the heart of it all: the concept of presence. J Comput Mediat Commun 2006;3(2). [doi: 10.1111/j.1083-6101.1997.tb00072.x]
34. Heeter C. Reflections on real presence by a virtual person. Presence (Camb) 2003;12(4):335-345. [doi: 10.1162/105474603322391587]
35. Slater M, Guger C, Edlinger G, Leeb R, Pfurtscheller G, Antley A, et al. Analysis of physiological responses to a social situation in an immersive virtual environment. Presence (Camb) 2006;15(5):553-569. [doi: 10.1162/pres.15.5.553]
36. Schachter M, Fins JJ. Informed consent revisited: a doctrine in the service of cancer care. Oncologist 2008 Oct;13(10):1109-1113 [FREE Full text] [doi: 10.1634/theoncologist.2008-0101] [Medline: 18849321]
37. del Carmen MG, Joffe S. Informed consent for medical treatment and research: a review. Oncologist 2005 Sep;10(8):636-641 [FREE Full text] [doi: 10.1634/theoncologist.10-8-636] [Medline: 16177288]
38. Festinger DS, Ratanadilok K, Marlowe DB, Dugosh KL, Patapis NS, Dematteo DS. Neuropsychological functioning and recall of research consent information among drug court clients. Ethics Behav 2007;17(2):163-186. [doi: 10.1080/10508420701378115]
39. Festinger DS, Dugosh KL, Croft JR, Arabia PL, Marlowe DB. Corrected feedback: a procedure to enhance recall of informed consent to research among substance abusing offenders. Ethics Behav 2010;20(5):387-399. [doi: 10.1080/10508422.2010.491767]
40. Palmer BW, Savla GN. The association of specific neuropsychological deficits with capacity to consent to research or treatment. J Int Neuropsychol Soc 2007 Nov;13(6):1047-1059. [doi: 10.1017/S1355617707071299] [Medline: 17942022]
41. Johnson-Greene D. Evolving standards for informed consent: is it time for an individualized and flexible approach? In: Bersoff DN, editor. Ethical Conflicts in Psychology, 4th edition. Washington, DC: American Psychological Association; 2008:337-338.
42. Gorini A, Gaggioli A, Vigna C, Riva G. A second life for eHealth: prospects for the use of 3-D virtual worlds in clinical psychology. J Med Internet Res 2008;10(3):e21 [FREE Full text] [doi: 10.2196/jmir.1029] [Medline: 18678557]

Abbreviations

BPVS: British Picture Vocabulary Scale
MAT: Matrix Analogies Test

Edited by G Eysenbach; submitted 30.08.11; peer-reviewed by J Keelan, C Thornson; comments to author 26.09.11; revised version received 17.10.11; accepted 19.10.11; published 14.11.11.

Please cite as:
Hall V, Conboy-Hill S, Taylor D
Using Virtual Reality to Provide Health Care Information to People With Intellectual Disabilities: Acceptability, Usability, and Potential Utility
J Med Internet Res 2011;13(4):e91
URL: http://www.jmir.org/2011/4/e91/
doi:10.2196/jmir.1917
PMID:22082765

XSL•FO
RenderX

©Valerie Hall, Suzanne Conboy-Hill, Dave Taylor. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development and Validation of Filters for the Retrieval of Studies of Clinical Examination From Medline

Nader Shaikh[1], MD MPH; Robert G Badgett[2], MD; Mina Pi[3], BS; Nancy L Wilczynski[4], PhD; K. Ann McKibbon[4], PhD; Andrea M Ketchum[5], MLIS; R. Brian Haynes[4], MD PhD

[1]University of Pittsburgh School of Medicine, General Academic Pediatrics, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA, United States

[2]Department of Internal Medicine, University of Kansas School of Medicine at Wichita, Wichita, KS, United States

[3]Department of Biology, Carnegie Mellon University, Pittsburgh, PA, United States

[4]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada

[5]Health Sciences Library System, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Nader Shaikh, MD MPH
University of Pittsburgh School of Medicine
General Academic Pediatrics
Children's Hospital of Pittsburgh of UPMC
4401 Penn Avenue
Pittsburgh, PA, 15224
United States
Phone: 1 412 692 8111
Fax: 1 412 692 8516
Email: nader.shaikh@chp.edu

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2012/4/e108

Abstract

Background: Efficiently finding clinical examination studies—studies that quantify the value of symptoms and signs in the diagnosis of disease—is becoming increasingly difficult. Filters developed to retrieve studies of diagnosis from Medline lack specificity because they also retrieve large numbers of studies on the diagnostic value of imaging and laboratory tests.

Objective: The objective was to develop filters for retrieving clinical examination studies from Medline.

Methods: We developed filters in a training dataset and validated them in a testing database. We created the training database by hand searching 161 journals (n = 52,636 studies). We evaluated the recall and precision of 65 candidate single-term filters in identifying studies that reported the sensitivity and specificity of symptoms or signs in the training database. To identify best combinations of these search terms, we used recursive partitioning. The best-performing filters in the training database as well as 13 previously developed filters were evaluated in a testing database (n = 431,120 studies). We also examined the impact of examining reference lists of included articles on recall.

Results: In the training database, the single-term filters with the highest recall (95%) and the highest precision (8.4%) were diagnosis[subheading] and "medical history taking"[MeSH], respectively. The multiple-term filter developed using recursive partitioning (the RP filter) had a recall of 100% and a precision of 89% in the training database. In the testing database, the Haynes-2004-Sensitive filter (recall 98%, precision 0.13%) and the RP filter (recall 89%, precision 0.52%) showed the best performance. The recall of these two filters increased to 99% and 94% respectively with review of the reference lists of the included articles.

Conclusions: Recursive partitioning appears to be a useful method of developing search filters. The empirical search filters proposed here can assist in the retrieval of clinical examination studies from Medline; however, because of the low precision of the search strategies, retrieving relevant studies remains challenging. Improving precision may require systematic changes in the tagging of articles by the National Library of Medicine.

XSL•FO
RenderX

(J Med Internet Res 2011;13(4):e82) doi:10.2196/jmir.1826

KEYWORDS

Medline; filter; hedge; clinical examination; recursive partitioning

Introduction

In arriving at a diagnosis, clinicians often rely on clinical examination findings (ie, information from the patient's history and/or physical examination) [1-3]. Therefore, easy availability of results from clinical examination studies can greatly influence medical care. The number of studies published per year that focus on clinical examination has more than tripled since 1980. As this literature multiplies, the task of reliably and simply identifying sound studies is becoming increasingly challenging.

In many areas of medicine, filters have been developed to facilitate the search for relevant articles. Filters are pretested search strategies that help identify studies of a certain type from among all the other studies in Medline. Search filters that are optimized for the retrieval of studies of diagnosis, therapy, and clinical prediction rules are available [4-6]. These filters are routinely used by both clinicians (eg, PubMed Clinical Queries [4]) and systematic reviewers (eg, Cochrane Highly Sensitive Search Strategy for therapy articles [7]). No published filters, however, have been developed to facilitate the search for studies of clinical examination [8]. A clinical examination filter may be useful to clinicians and authors of systematic reviews. Clinicians need to identify sound clinical examination articles in a timely fashion so that they can effectively care for their patients. With the commencement of Cochrane reviews of Diagnostic Test Accuracy [9], which will include reviews of clinical examination, there is a growing need for filters optimized for the retrieval of clinical examination studies.

The goal of this study was to develop and evaluate Medline filters that could facilitate retrieval of clinical examination studies.

Methods

Overview

The training and testing of the filters entailed 8 steps: (1) development of a training database, (2) identification of candidate single-term filters, (3) identification of single-term filters with the best performance in the training database, (4) identification of the multiple-term filter with the best performance in the training database using recursive partitioning, (5) development of a testing database, (6) evaluation of the performance of filters developed in this study in the testing database, (7) evaluation of the performance of previously developed filters in the testing database, and (8) examination of the impact of reviewing reference lists of included articles on recall. We performed our research using PubMed, the United States National Library of Medicine's public search engine for accessing Medline.

Development of a Training Database

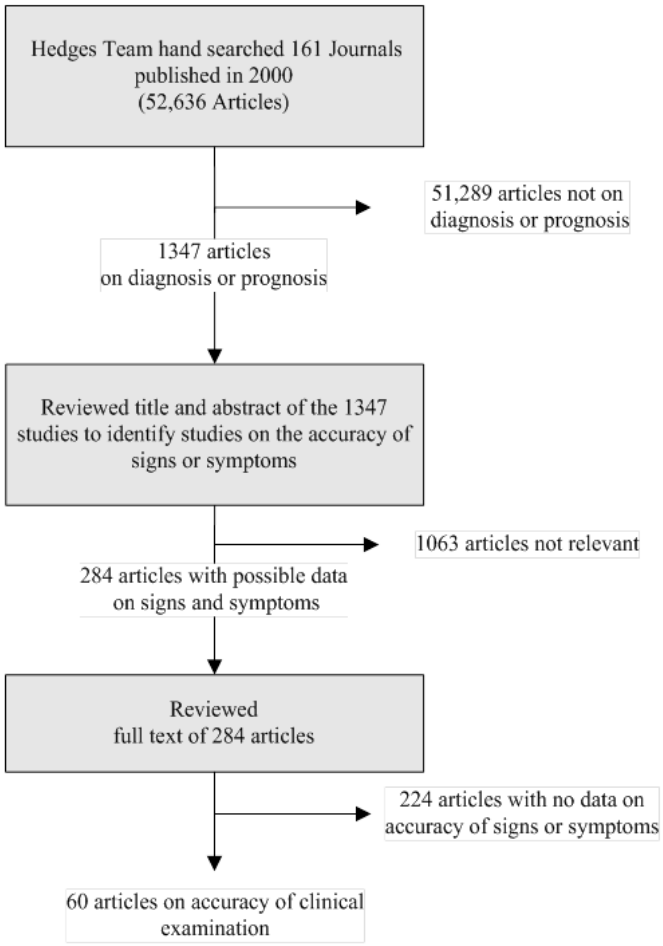
We used the Clinical Hedges database, the methods of which have been previously described [10], as the starting point for this study. Briefly, the Hedges team conducted a hand search of articles published in the year 2000 in 161 prominent journals that met criteria for high quality; a total of 52,636 articles were reviewed. The team categorized articles as pertaining to diagnosis, therapy, or prognosis (among other categories) based on a priori criteria. For the project reported here, we reviewed the studies identified in the Clinical Hedges database as pertaining to diagnosis or prognosis to identify those that specifically pertained to clinical examination (n = 1347).

One investigator (author NS) initially reviewed the title and abstract (if an abstract was available) and full text, if necessary, of the 1347 studies and classified each article as a *clinical examination* (gold standard article) or a *non–clinical examination* article (Figure 1). Gold standard articles were those that met our a priori criteria for quantifying the value of the clinical examination. We only considered physical examination findings that could be elicited with commonly available props such as a stethoscope or ophthalmoscope. We included articles that reported both sensitivity and specificity for at least one symptom, sign, or a combination of signs and symptoms (Figure 1). We included multivariable diagnostic rules if they were composed of only signs or symptoms; studies describing a multivariable rule that included imaging or laboratory findings were not considered because these studies can easily be found using existing, more general filters designed for the detection of diagnostic tests. For example, the Breese score—a validated scoring system to diagnose streptococcal pharyngitis in children—was not considered a clinical examination study because, in addition to signs and symptoms, a white blood cell count is required to calculate a total score [11]. We excluded studies of prognostic factors, that is, those focusing on the prediction of future disease (eg, prediction of mortality based on findings on admission to the intensive care unit). Articles with less than 10 patients were excluded because these studies, due to their very small sample size, cannot provide accurate estimates of sensitivity or specificity. Studies that could not be easily categorized were independently reviewed by a second reviewer (author RGB) and differences were resolved by discussion. This process resulted in 60 of 52,636 articles meeting the gold standard criteria (Figure 1).

We then recreated the Clinical Hedges dataset by entering the 161 journals in Medline and by restricting the publication year to 2000 (Figure 1). We placed the articles into two collections stored in an account at PubMed. One collection contained the articles that met our criteria for gold standard, and the other collection contained the remaining articles.

XSL•FO
RenderX

Figure 1. Flow sheet describing development of the training database.



Identification of Candidate Single-Term Filters

We generated a list of 65 candidate search terms in PubMed syntax with the help of two clinicians, three reference librarians, and a thorough review of the literature. The expert searchers independently reviewed our lists of candidate terms and suggested additional terms. We used terms pertaining to clinical examination and diagnosis as well as negated terms (eg, NOT MRI). (See Multimedia Appendix 1 for a list of the search terms used.) The following PubMed syntax was used: [tw] = text word; [MeSH] = National Library of Medicine's Medical Subject Heading; [sh] = MeSH subheading; [TIAB] = Title or abstract; [pt] = publication type; [ti] = title; du[sh] = diagnostic use MeSH subheading; noexp = do not explode (ie, do not automatically include the more specific terms beneath the MeSH term in the MeSH hierarchy).

Identification of Best-Performing Single Term Filters Using the Training Database

We evaluated each individual filter against the training database to determine its recall (proportion of the clinical examination articles that the filter detected), precision (proportion of articles retrieved that were relevant), F-measure (an overall measure combining recall and precision), "fallout" (the proportion of nonrelevant articles that were retrieved), and the number needed to read (the average number of articles the searcher will need to look at to find each relevant article) [12]. Of the clinical examination terms, 7 had a recall of greater than 25% and a fallout of less than 50%. We evaluated all possible combinations (2-term combinations, 3-term combinations, 4-term combinations, 5-term combinations, 6-term combinations, and one 7-term combination) of these 7 terms to identify the combinations with the highest recall, precision, and F-measure. We repeated this process for all possible combination of the 8 diagnosis terms with a recall of greater than 25% and a fallout of less than 50%.

Development of a Multiple-Term Filter Using Recursive Partitioning Using the Training Database

Because testing all combinations of single-term filters would have been prohibitive, we used recursive partitioning to develop the best multiple term filter (hereinafter referred to as the recursive partitioning filter) [13]. Recursive partitioning is a form of nonparametric discriminant analysis that repeatedly stratifies the group into smaller mutually exclusive subgroups according to a set of predictor variables. Apart from its efficiency, an added advantage of recursive partitioning is its ability to create filters including both Boolean terms, OR and AND. Recursive partitioning also adds the ability to vary misclassification costs (cost of a false positive vs costs of a false negative) in order to identify terms that best address the objectives of the analysis. For each of the 41 terms with a recall of greater than 25% *or* fallout less than 75%, we calculated the recall, precision, F-measure, and fallout against the training

XSL•FO
RenderX

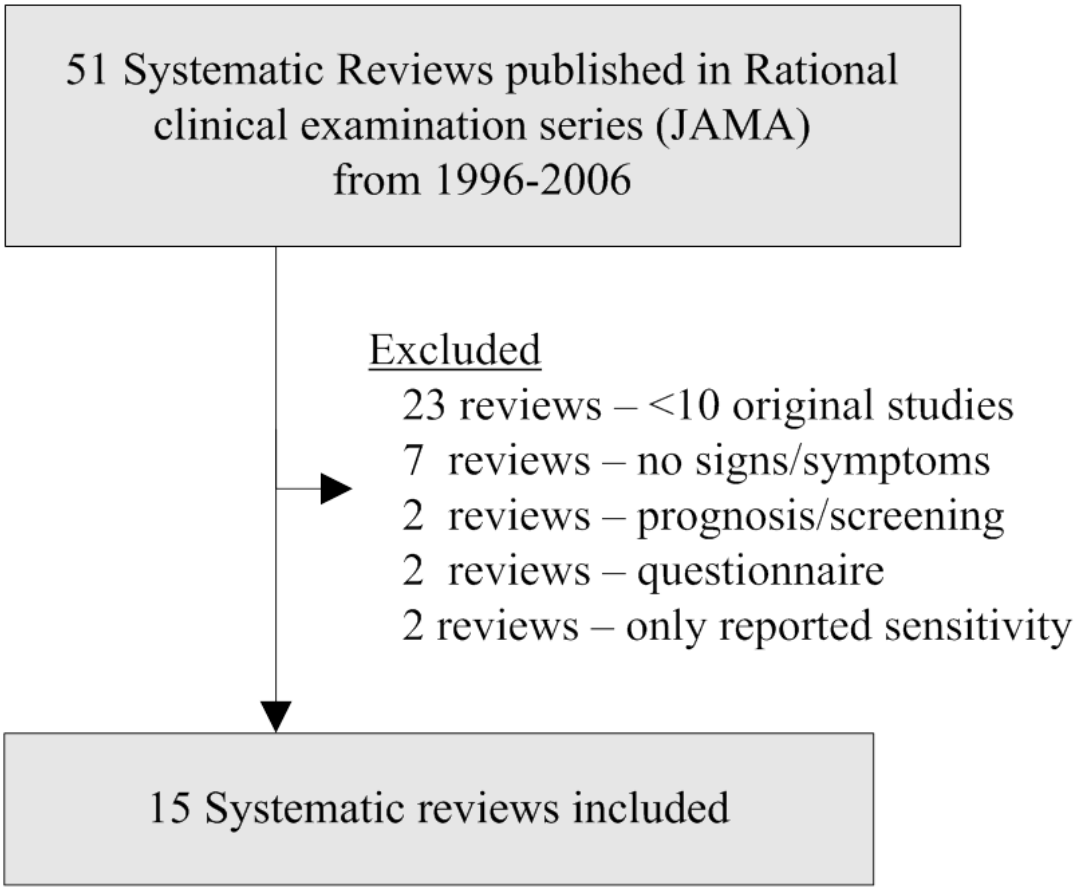
database. To decide on the first branching point in the tree, we chose the term with the lowest weighted error rate (weight based on the prevalence of clinical examination studies among all studies in the database) [13]. Once the term with the lowest error rate was found (ie, diagnosis[tw]), we created 4 new datasets in PubMed ("clinical examination" AND diagnosis[tw]; "clinical exam" NOT diagnosis[tw]; "non clinical examination" AND diagnosis[tw]; "Non clinical exam" NOT diagnosis[tw]). We then tested all remaining filters against each of these 4 new datasets and again identified terms with the lowest error rate. This allowed us to grow the recursive partitioning tree. We repeated this until the two 2x2 tables created by the split were no longer significantly different from each other ($P > .05$). Because this approach can lead to overfitting, we also required each new branch to have a recall of at least 99%.

Establishing the Testing Database

To develop the testing database, we used the largest collection of systematic reviews on clinical examination in the literature: The Rational Clinical Examination series in the Journal of the American Medical Association (JAMA) [14]. One author (NS) used a priori inclusion and exclusion criteria similar to those used to establish the training database to develop the testing database. We included systematic reviews that reviewed at least 10 original studies, reported sensitivity and specificity of signs or symptoms, and were published beginning in 1996 through 2006 (Figure 2). Reviews that did not pertain to clinical examination, reviews of questionnaires, reviews of multivariable diagnostic rules that included laboratory or imaging variables, and reviews of prognostic or screening tests were excluded. A total of 15 systematic reviews met all inclusion criteria.

Articles included in these 15 reviews were regarded as *relevant* articles (gold standard) that the filters should be able to recall. To identify *nonrelevant* articles we recreated the subject-specific search (eg, temporal arteritis or giant cell arteritis) using the search strategy reported in the methods section of each of the systematic reviews; articles that were retrieved by the electronic subject-specific search but that were not included in the review were regarded as nonrelevant. This allowed us to calculate the number of relevant and nonrelevant articles for each review. A total of 224 original clinical examination articles were included in these 15 systematic reviews. We excluded 7 articles that were not in Medline. One study was excluded because it was not found by the subject-specific search. In all, 28 older studies without abstracts were excluded because filters would have difficulty retrieving these studies and because contemporary studies of the clinical examination are likely to have abstracts. The resulting testing database included 188 articles that were relevant and 430,932 articles that were nonrelevant.

Figure 2. Flow sheet describing development of the testing database.



XSL•FO
RenderX

Evaluation of the Filters Developed in This Study

For the 3 filters with the highest recall in the training database, we calculated the recall, precision, F-measure, and the number needed to read in the testing database. The calculations were based on the cells and formulas in Table 1.

Table 1. A 2x2 table created for each systematic review and formulas used[a]

	Articles Included in the Systematic Review	Articles Not Included in the Systematic Review
Detected by filter	A	B
Missed by filter	C	D

[a] Recall = A/(A+C); Precision = A/(A+B); F-measure = 2*precision*recall/(precision + recall); Number needed to read = 1/precision; Fallout = B/(B+D) [15,16]

Evaluation of Previously Developed Filters

The performance of 12 previously developed filters validated for retrieving articles on diagnosis [10,17-21] and 1 filter developed specifically for the clinical examination by editors of the Rational Clinical Examination series [22] was evaluated in the testing database. The filters tested are listed in Multimedia Appendix 2 and are named using the name of the first author of the publication describing the filter followed by the year of publication. If more than one filter was described in the publication, we tagged on the name used by the author to describe the various filters. For example, the label "Haynes-2004-Sensitive" refers to the filter described by Haynes et al in their 2004 publication with the highest sensitivity (ie, the filter with the highest recall). Finally, we tested whether a combination of the best filters would improve performance.

Impact of Reviewing Reference Lists on Recall

Authors of systematic reviews often examine reference lists hoping to increase recall. We examined how this strategy would complement the use of filters in the area of clinical examination. Specifically, we examined whether checking the reference lists of included articles would allow use of a filter with a lower recall. Thus, we identified articles that were missed by the 2 filters with the highest recall and checked to see if these articles were included in the reference lists of the articles not missed by these filters.

Results

Training Results

Filters with the best performance in the training database are shown in Table 2. The term diagnosis[subheading] identified 95% of the clinical examination studies. The MeSH term physical examination identified only 25% of studies and was therefore not included in the table. In general, multiple-term search filters using only terms pertaining to diagnosis outperformed the filters using only clinical examination terms. Also, 3 filters had a recall of 100% (CE-high recall, Dx-high recall, RP) and two of these (Dx-high recall, RP) appeared particularly promising because of their higher precision.

XSL•FO
RenderX

Table 2. Filters with the best recall (keeping fallout less than 50%), precision (keeping recall greater than 50%) and F-measure in the training database

Filter	Performance Measure	Recall (%)	Precision (%)	F-measure	NNR[a]
Best single-term filter					
Diagnosis[subheading]	Best recall	95	0.35	0.71	279
Medical history taking[MeSH]	Best precision and F-measure	12	8.44	9.79	11.86
Best multiple-term filters using only diagnosis terms					
Diagnosis[tw] OR "sensitivity and specificity"[MeSH]	Best recall (hereinafter Dx-high recall)	100	0.52	1.04	191
Predictive value of tests[mesh] OR specificity[TIAB]	Best precision and F-measure (hereinafter Dx-precise)	67	1.95	3.78	51
Best multiple-term filters using only clinical examination terms					
Clinical*[tw] OR symptom*[tw] OR exam*[tw] OR criteria[tw] OR tests[tw] OR test[tw]	Best recall (hereinafter CE-high recall)	100	0.27	0.53	377
Tests[tw] OR physical[tw]	Best precision and F-measure (hereinafter CE-precise)	62	0.72	1.43	138
Best multiple-term filter using all terms					
(Diagnosis[tw] AND (specific*[tw] OR clinical*[tw] OR exam*[tw])) OR "sensitivity and specificity"[MeSH]	Best overall filter from recursive partition (hereinafter RP-filter)[b]	100	0.89	1.76	113

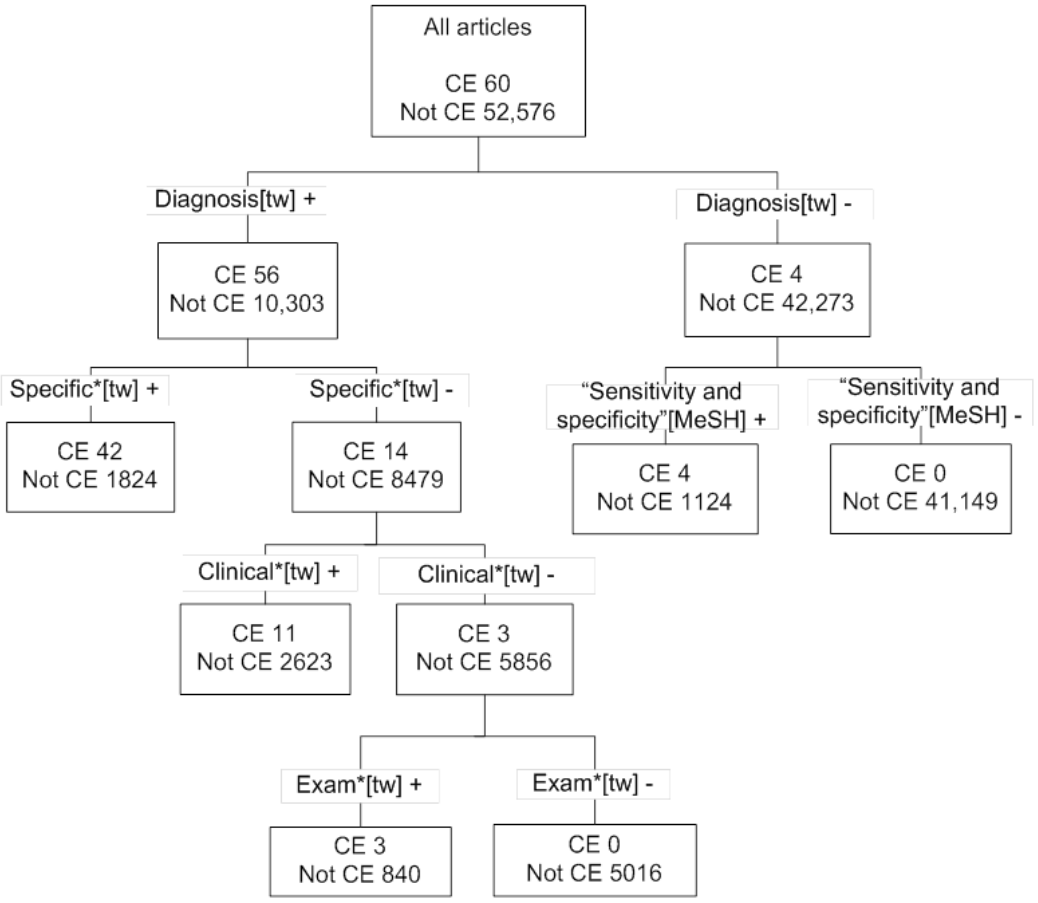
[a] Number needed to read

[b]Filter developed using recursive partitioning (see “Methods” section)

The recursive partitioning tree is shown in Figure 3. When converted to Boolean language, the RP filter, in PubMed syntax, is as follows: (Diagnosis[tw] AND (specific*[tw] OR clinical*[tw] OR exam*[tw])) OR "sensitivity and specificity"[MeSH].

XSL•FO
RenderX

Figure 3. Best multiple-term filter for retrieval of articles on clinical examination (CE) developed using recursive partitioning.



Testing Results

The recall, precision, F-measure, and the number needed to read for the filters developed in this study as well as the 13 previously developed filters and combination of filters are presented in Table 3. The Haynes-2004-Sensitive filter[10] had the highest recall (98%). When considering only filters with a recall of 80%, the RP filter had the highest precision (0.26%). The Haynes-2004-Sensitive filter and the CE-high recall filter when combined using the Boolean term OR had a recall of 100% and a precision of 0.06%. Other filter combinations did not offer much of an improvement in recall compared with their individual use.

XSL•FO
RenderX

Table 3. Performance of the search filters in the testing database sorted according to recall

Filters or Filter Combinations	Recall (%)	Precision (%)	F-measure	NNR[a]
Filters				
Haynes-2004-Sensitive [10]	98	0.13	0.26	778
Vincent-2003 [21]	98	0.09	0.17	1154
Bachmann-2002 [15]	96	0.11	0.22	906
Haynes-1994-Sensitive [19]	95	0.16	0.31	641
Dx-high recall[b]	95	0.12	0.25	804
Van der Weijden-1997 [20]	95	0.07	0.13	1490
CE-high recall[b]	91	0.08	0.15	1330
Haynes-1994-Accurate [19]	91	0.07	0.14	1431
RP-filter[b]	89	0.26	0.52	380
Rational Clinical exam [22]	73	0.30	0.61	328
Deville-2002 [18]	71	0.40	0.80	249
Haynes-2004-Accurate [10]	69	0.45	0.89	224
Deville-2000-Accurate [17]	64	0.64	1.26	157
Deville-2000-Sensitive [17]	64	0.60	1.19	167
Haynes-1994-Specific [19]	51	0.72	1.42	139
Haynes-2004-Specific [10]	36	1.01	1.97	99
Filter combinations				
Haynes-2004-Sensitive [10] OR CE-high recall	100	0.06	0.12	1613
CE-high recall OR RP	99	0.06	0.13	1572
Haynes-2004-Sensitive [10] OR RP	98	0.11	0.22	890
Haynes-2004-Sensitive [10] AND RP	95	0.13	0.25	790
Haynes-2004-Sensitive [10] AND CE-high recall	88	0.19	0.39	515

[a]NNR = number needed to read

[b]The three filters with highest recall in the training database

Impact of Reviewing Reference Lists

Overall, 4 of 188 relevant articles were missed by the Haynes-2004-Sensitive search strategy, and, of these, 2 were retrieved by reviewing the reference lists of the articles not missed by this strategy (increasing recall from 98% to 99%). Of the 19 articles missed by the recursive partitioning strategy, 8 were retrieved by reviewing the reference lists of the articles not missed by this strategy (increasing recall from 89% to 94%).

Discussion

We quantified the recall and precision of filters that may be used to find articles on clinical examination in MEDLINE. While the use of recursive partitioning may increase the precision of searching, all of the strategies we tested had a very low precision of less than 2%.

Application of the Filters

For health care providers looking for information regarding the diagnostic accuracy of clinical examination findings, the RP filter appears to be the most reasonable choice. For example, let us assume that a clinician is reviewing the ability of the third heart sound to detect heart failure. To determine the posttest probability of congestive heart failure among patients with a third heart sound, the search using the RP filter in PubMed would be (gallop OR S3 OR third heart sound) AND heart failure[MeSH] AND ((Diagnosis[tw] AND (specific*[tw] OR clinical*[tw] OR exam*[tw])) OR "sensitivity and specificity"[MeSH]). As of March 2011 this search yielded 68 articles, several of which directly related to the clinician's question. Although not studied, the physician could restrict the search to systematic reviews by adding the term "systematic[sb]". This strategy yielded 1 relevant systematic review. While the NNRs for the filters examined reported in this study are very high (Table 3), the NNR will be considerably lower in clinical practice. The NNR, like the positive predictive value of a diagnostic test, is dependent on the prevalence of articles about physical examination. Although the proportion of physical examination studies in MEDLINE is relatively low (eg, < 0.1% in the Hedges database), when the clinician enters search terms for a disease and for the physical examination findings, the prevalence of physical examination articles

XSL•FO
RenderX

increases. As a result, the number needed to read will be substantially lower (see example above). Accordingly, it is critical that the clinician uses a well-built clinical question using the most descriptive and specific terms possible [23].

For the researcher who wants to undertake a systematic review, the Haynes-2004-Sensitive filter [10], with its 98% recall, appears to be the most reasonable choice. Nevertheless, some articles may be missed if one relies on this filter alone. Two strategies are suggested for increasing recall. One is to examine the reference list of the articles meeting criteria for inclusion. This increases the sensitivity to 99%. The other strategy is to combine the Haynes-1994-and the CE High recall filter using OR. Although this strategy had a 100% recall, its precision was very low (0.06%). Even though relying on filters alone may lead to some studies being missed [24], we feel that use of filters is appropriate, especially when it is exceedingly difficult to conduct a review without one. The filters presented here are intended to be used as part of a larger search strategy, which includes a review of reference lists, and communication with experts in the field.

Poor Precision of Filters for Clinical Examination Studies

All of the filters we tested had a very low precision in identifying clinical examination studies. Our findings are consistent with findings published by Haynes and colleagues [19], indicating poor precision of filters developed for retrieving articles on diagnosis as compared to those developed for retrieving articles on treatment (Table 4). These observations suggest that the National Library of Medicine should create a publication type for studies that quantify sensitivity and specificity for diagnosis. Other alternative or complementary solutions may involve manually identifying and tagging studies that quantify the clinical examination (as is currently used by the Cochrane Collaboration to create a database of sound randomized controlled treatment trials, CENTRAL), collaborative filtering, or content-based filtering [25].

Table 4. Comparison of the performance of filters for clinical examination, diagnosis, and treatment

Filters	Recall (%)	Precision (%)	F-measure	NNR[a]
Clinical examination				
Haynes-2004-Sensitive [10]	98	0.13	0.26	778
Recursive partitioning	89	0.26	0.52	380
Diagnosis in general				
Haynes-2004-Sensitive [10]	99	1.1	2.17	91
Treatment				
Haynes 2005 [26][b]	99	9.9	18.0	10
Haynes 1994 [19][b]	99	22	36.0	4.5

[a]NNR = Number needed to read

[b]Values are for the most-sensitive multi-term filter

Limitations

There are several limitations to our study. The Hedges database [10] contains the 161 journals whose articles were felt to have the highest scientific merit and clinical relevance. While we believe these are the journals that will most help clinicians, the results may vary when all of Medline is searched. In addition, journals published in foreign languages were not included in the Hedges database. Because some of our filters used text words, these filters may fall in performance when searching for articles that have not been translated to English or for articles without an abstract. Another limitation was in our identification of candidate search terms. Consistent with prior studies of filter development, expert searchers independently reviewed our lists of candidate terms and suggested additional terms. However, we did not quantitatively review the most frequent search terms and text words in the gold standard studies to identify candidate terms. However, when we retrospectively examined the MeSH terms that were used to index the gold standard studies in the training database, the terms not tested by us had substantially lower recall and precision than the terms we selected. Nevertheless, we believe that future studies should incorporate this method of identifying terms. Another limitation was in identification of the gold standard articles in the training database. Only one investigator initially reviewed the articles for eligibility. Future studies should utilize two investigators who independently assess each article. Finally, because of the low prevalence of clinical examination studies, the number of gold standard studies in both the training and testing databases were relatively small. Further testing of these filters in larger databases is necessary.

A surprising result is that only 25% and 20% of the clinical examination studies in the training database were coded with the MeSH terms “physical examination” and “signs and symptoms”, respectively. This current inconsistency in the assignment of these MeSH terms limits the ability of search filters on this topic.

Implications for Future Filter Development

We present a new method for the development of multi-term filters. The use of recursive partitioning in the development of filters is novel and seems particularly well suited when there are many candidate terms. When the number of candidate terms is small, one could test all the possible combinations of terms against the dataset. This becomes prohibitive when the number

XSL•FO
RenderX

of candidate terms is large. In contrast, using recursive partitioning, a search filter is constructed in a stepwise fashion. This method also allows for the development of filters that use both AND and OR terms and allows for the development of filters with the best combination of recall and precision.

Conclusions

Recursive partitioning offers an alternative method of developing filters: it not only allows for the development of filters with the best combination of recall and precision, but also for the development of filters that use both AND and OR Boolean connectors. Despite the advantages of recursive partitioning, the filters we developed for the retrieval of clinical examination studies had relatively low precision. We believe the National Library of Medicine should create a publication type for articles that quantify the sensitivity and specificity of the clinical examination. This new tag could improve retrieval of studies of clinical diagnosis.

Acknowledgments

We acknowledge Chen-Pin Wang, PhD, Department of Epidemiology and Biostatistics at the University of Texas Health Science Center at San Antonio, who helped us develop statistical strategies to address overfitting in the recursive partitioning model.

Conflicts of Interest

None declared

Multimedia Appendix 1

List of single-term filters.

[PDF File (Adobe PDF File), 76KB - jmir_v13i4e82_app1.pdf]

Multimedia Appendix 2

List of filters evaluated in the testing corpus.

[PDF File (Adobe PDF File), 61KB - jmir_v13i4e82_app2.pdf]

References

1. Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. Br Med J 1975 May 31;2(5969):486-489. [Medline: 1148666]
2. Peterson MC, Holbrook JH, Von Hales D, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. West J Med 1992 Feb;156(2):163-165. [Medline: 1536065]
3. Sandler G. Costs of unnecessary tests. Br Med J 1979 Jul 7;2(6181):21-24. [Medline: 466256]
4. Haynes RB, Wilczynski N. Finding the gold in MEDLINE: clinical queries. ACP J Club 2005;142(1):A8-A9. [Medline: 15656539]
5. Wilczynski NL, Haynes RB, Hedges Team. EMBASE search strategies achieved high sensitivity and specificity for retrieving methodologically sound systematic reviews. J Clin Epidemiol 2007 Jan;60(1):29-33. [doi: 10.1016/j.jclinepi.2006.04.001] [Medline: 17161751]
6. Wong SS, Wilczynski NL, Haynes RB, Ramkissoonsingh R, Hedges Team. Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. AMIA Annu Symp Proc 2003:728-732. [Medline: 14728269]
7. Lefebvre C, Manheimer E, Glanville J. Cochrane Handbook for Systematic Reviews of Interventions. Oxfordshire, UK: The Cochrane Collaboration; 2011. Searching for studies URL: http://www.mrc-bsu.cam.ac.uk/cochrane/handbook/index.htm [accessed 2011-10-04] [WebCite Cache ID 62BUu8FCa]
8. Simel DL, Rennie D, Bossuyt PM. The STARD statement for reporting diagnostic accuracy studies: application to the history and physical examination. J Gen Intern Med 2008 Jun;23(6):768-774. [doi: 10.1007/s11606-008-0583-3] [Medline: 18347878]
9. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008 Dec 16;149(12):889-897. [Medline: 19075208]
10. Haynes RB, Wilczynski NL. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. BMJ 2004 May 1;328(7447):1040 [FREE Full text] [doi: 10.1136/bmj.38068.557998.EE] [Medline: 15073027]
11. Breese BB. A simple scorecard for the tentative diagnosis of streptococcal pharyngitis. Am J Dis Child 1977 May;131(5):514-517. [Medline: 855837]
12. Hersh WR. Information Retrieval: A Health Care Perspective. New York: Springer; 1995.

XSL•FO
RenderX

13. Cook EF, Goldman L. Empiric comparison of multivariate analytic techniques: advantages and disadvantages of recursive partitioning analysis. J Chronic Dis 1984;37(9-10):721-731. [Medline: 6501544]
14. Simel DL, Rennie D. The clinical examination. An agenda to make it more rational. JAMA 1997 Feb 19;277(7):572-574. [Medline: 9032165]
15. Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. J Am Med Inform Assoc 2002;9(6):653-658 [FREE Full text] [Medline: 12386115]
16. Toth B, Gray JA, Brice A. The number needed to read-a new measure of journal value. Health Info Libr J 2005 Jun;22(2):81-82. [doi: 10.1111/j.1471-1842.2005.00568.x] [Medline: 15910578]
17. Devillé WL, Bezemer PD, Bouter LM. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. J Clin Epidemiol 2000 Jan;53(1):65-69. [Medline: 10693905]
18. Devillé WL, Buntinx F, Bouter LM, Montori VM, de Vet HC, van der Windt DA, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol 2002 Jul 3;2:9 [FREE Full text] [Medline: 12097142]
19. Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. J Am Med Inform Assoc 1994;1(6):447-458 [FREE Full text] [Medline: 7850570]
20. van der Weijden T, IJzermans CJ, Dinant GJ, van Duijn NP, de Vet R, Buntinx F. Identifying relevant diagnostic studies in MEDLINE. The diagnostic value of the erythrocyte sedimentation rate (ESR) and dipstick as an example. Fam Pract 1997 Jun;14(3):204-208 [FREE Full text] [Medline: 9201493]
21. Vincent S, Greenley S, Beaven O. Clinical Evidence diagnosis: Developing a sensitive search strategy to retrieve diagnostic studies on deep vein thrombosis: a pragmatic approach. Health Info Libr J 2003 Sep;20(3):150-159. [Medline: 12919278]
22. Society of General Internal Medicine. Rational Clinical Examination Series. Alexandria, VA: Society of General Internal Medicine; 2008. Instructions for authors for rational clinical examination URL: http://sgim.org/index.cfm?pageId=666 [accessed 2011-04-12] [WebCite Cache ID 5xtMO85OU]
23. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007;7:16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
24. Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin Epidemiol 2006 Mar;59(3):234-240. [doi: 10.1016/j.jclinepi.2005.07.014] [Medline: 16488353]
25. Yoneya T, Mamitsuka H. PURE: a PubMed article recommendation system based on content-based filtering. Genome Inform 2007;18:267-276 [FREE Full text] [Medline: 18546494]
26. Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR, Hedges Team. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. BMJ 2005 May 21;330(7501):1179 [FREE Full text] [doi: 10.1136/bmj.38446.498542.8F] [Medline: 15894554]

Edited by G Eysenbach; submitted 13.04.11; peer-reviewed by L Zhang, E Leclercq; comments to author 09.05.11; revised version received 03.06.11; accepted 06.06.11; published 19.10.11.

Please cite as:
Shaikh N, Badgett RG, Pi M, Wilczynski NL, McKibbon KA, Ketchum AM, Haynes RB
Development and Validation of Filters for the Retrieval of Studies of Clinical Examination From Medline
J Med Internet Res 2011;13(4):e82
URL: http://www.jmir.org/2011/4/e82/
doi:10.2196/jmir.1826
PMID:22011384

©Nader Shaikh, Robert G. Badgett, Mina Pi, Nancy L. Wilczynski, K. Ann McKibbon, Andrea M. Ketchum, R. Brian Haynes. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 19.10.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Do Family Physicians Retrieve Synopses of Clinical Research Previously Read as Email Alerts?

Roland Grad[1], MD MSc; Pierre Pluye[1], MD PhD; Janique Johnson-Lafleur[1], MSc; Vera Granikov[1], MLIS; Michael Shulha[1], MLIS; Gillian Bartlett[1], PhD; Bernard Marlow[2], MD

[1]Information Technology Primary Care Research Group, Department of Family Medicine, McGill University, Montreal, QC, Canada

[2]College of Family Physicians of Canada, Mississauga, ON, Canada

Corresponding Author:
Roland Grad, MD MSc
Information Technology Primary Care Research Group
Department of Family Medicine
McGill University
3755 Cote Ste Catherine Road
Montreal, QC, H3T1E2
Canada
Phone: 1 514 340 8222 ext 5851
Fax: 1 514 340 8300
Email: roland.grad@mcgill.ca

Abstract

Background: A synopsis of new clinical research highlights important aspects of one study in a brief structured format. When delivered as email alerts, synopses enable clinicians to become aware of new developments relevant for practice. Once read, a synopsis can become a *known item* of clinical information. In time-pressured situations, remembering a known item may facilitate information retrieval by the clinician. However, exactly how synopses first delivered as email alerts influence retrieval at some later time is not known.

Objectives: We examined searches for clinical information in which a synopsis previously read as an email alert was retrieved (defined as a *dyad*). Our study objectives were to (1) examine whether family physicians retrieved synopses they previously read as email alerts and then to (2) explore whether family physicians purposefully retrieved these synopses.

Methods: We conducted a mixed-methods study in which a qualitative multiple case study explored the retrieval of email alerts within a prospective longitudinal cohort of practicing family physicians. Reading of research-based synopses was tracked in two contexts: (1) push, meaning to read on email and (2) pull, meaning to read after retrieval from one electronic knowledge resource. Dyads, defined as synopses first read as email alerts and subsequently retrieved in a search of a knowledge resource, were prospectively identified. Participants were interviewed about all of their dyads. Outcomes were the total number of dyads and their type.

Results: Over a period of 341 days, 194 unique synopses delivered to 41 participants resulted in 4937 synopsis readings. In all, 1205 synopses were retrieved over an average of 320 days. Of the 1205 retrieved synopses, 21 (1.7%) were dyads made by 17 family physicians. Of the 1205 retrieved synopses, 6 (0.5%) were known item type dyads. However, dyads also occurred serendipitously.

Conclusion: In the single knowledge resource we studied, email alerts containing research-based synopses were rarely retrieved. Our findings help us to better understand the effect of push on pull and to improve the integration of research-based information within electronic resources for clinicians.

(J Med Internet Res 2011;13(4):e101) doi:10.2196/jmir.1683

KEYWORDS

Electronic mail; clinical email channels; information retrieval; physicians, family

XSL•FO
RenderX

Introduction

The environment of primary care medicine severely limits time for searches of clinical information. At the point of care, and given the time required for searches, using electronic knowledge resources during the consultation is perceived to be a complex task [1]. Away from the point of care, keeping up with the literature involves selecting and interpreting relevant clinical research, which is far from trivial.

Reading synopses of new clinical research delivered as email alerts allows clinicians to become aware of new developments relevant for practice [2-4]. A synopsis consists of important aspects of a research study presented in a brief structured format that allows for quick reading (see Figure 1). These synopses are often emailed on a daily or weekly basis [5]. To facilitate retrieval of synopses first delivered as email alerts, some electronic knowledge resources make these synopses available within searchable databases [6]. One example of such integration is Essential Evidence Plus featuring POEMs (patient-oriented evidence that matters) [7].

Our literature review of email alerts in clinical practice found only five evaluation studies in the health sciences [8-12]. Citation tracking of these papers and a subsequent literature search resulted in one study. This study demonstrated that email to adults from rural counties containing short updates of new content on a nutrition website increased usage of that website [13]. Outside of medicine, marketing research and business literature have long noted the ability of targeted and personalized email to increase traffic to websites, increase sales and revenue, and create an interactive relationship with the recipient [14-16]. In information science, the concept of known items and known-item searching has been explored since the early 1980s. It has been demonstrated that users of online library catalogs are more likely to be successful when searching for a known item as opposed to a more general subject search [17].

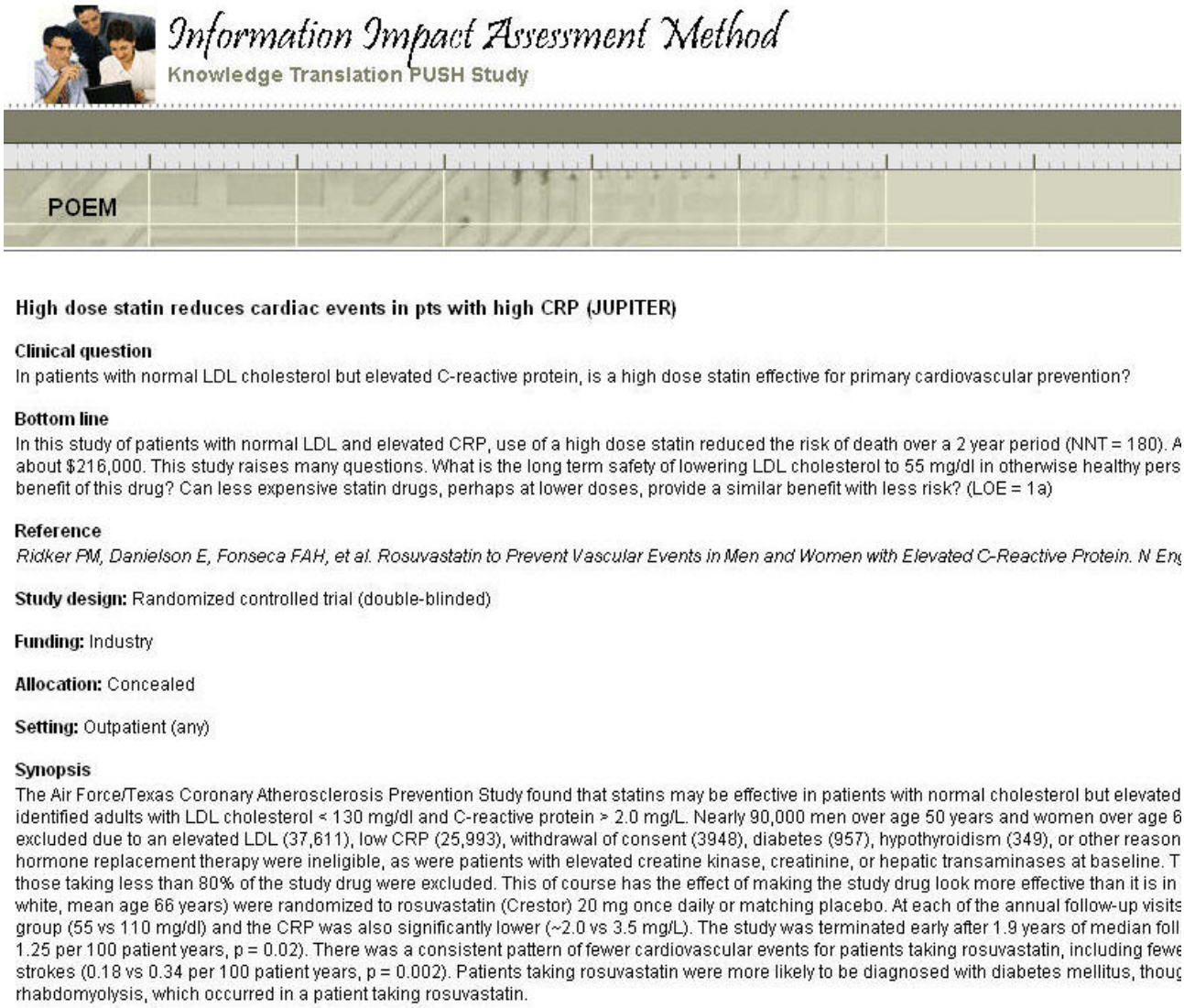
Previously, we have proposed a "push-pull" conceptual framework [18]. In this framework, it is assumed the *push* of clinical information will stimulate *pull* through the retrieval of objects of pushed information. In medicine, one study has examined the effect of push on pull [10]. In a cluster randomized trial of McMaster PLUS software, 203 physicians used either a full-serve version (that included email alerts to new articles and a cumulative database of email alerts) or a self-serve version that included the database and a passive guide to evidence-based literature. On average, physicians receiving the full-serve version made 0.77 more log-ins per month. How email alerts modestly increased log-ins to McMaster PLUS software was not reported. Thus, we do not know how push may influence pull in terms of retrieval of objects of pushed information.

Given the demands of practice and the limits of human memory, we assumed clinicians would occasionally need to retrieve information they had previously read as an email alert. Once read, email alerts can become known items of information. A search for a known item may include the author, the title, the subject, or a combination of these and other information [19]. If the push of synopses led to the creation of known items, retrieval of this information would be facilitated, helping to meet the demands of clinical practice in time-pressured situations. In addition, knowing about a synopsis might overcome one of the most common reasons given by physicians for not pursuing a clinical question—doubt about the existence of relevant information [20]. Therefore, we conducted a study of how the push of synopses of clinical research can lead to their subsequent retrieval by family physicians. We did this by prospectively identifying push-pull events operationalized as dyads. A dyad was defined as an occurrence of a family physician retrieving a synopsis from a knowledge resource when that synopsis had been read previously as an email alert.

Our study objectives were to (1) examine the retrieval of synopses from a knowledge resource among family physicians reading synopses as email alerts and then (2) using brief interviews, explore whether family physicians purposefully retrieved synopses that had been previously read as email alerts.

XSL•FO
RenderX

Figure 1. Example of a synopsis.



Methods

Study Design and Participants

A mixed-methods study was conducted using a *validation* design [21]. A qualitative multiple case study explored results from a prospective longitudinal cohort. We chose an exploratory naturalistic approach given that we did not know either the frequency or the variety of reasons why physicians retrieve synopses they previously read as email alerts. From 9 of 10 provinces, 41 family physicians consented to participate. Of these, 36 were certificants of the College of Family Physicians of Canada (CFPC). There were 24 men and 17 women ranging in age from 28 to 70 years (median 44 years). In addition, of these 41 family physicians, 28 (68%) had a faculty appointment, and all were in active practice. With respect to their main patient setting, 1 family physician had no Internet access, 37 (90%) reported having high-speed access, and 3 did not know what type of connection they had. In terms of computer self-efficacy, 8 (20%) rated their level of skill as advanced, 32 (78%) as intermediate, and 1 as beginner. Early on, 1 participant dropped out of the study before retrieving any synopses. The study protocol was approved by the McGill University Faculty of Medicine Institutional Review Board.

Quantitative Methods With Respect to Objective 1: Do Family Physicians Retrieve Synopses They Previously Read as Email Alerts?

Data Collection

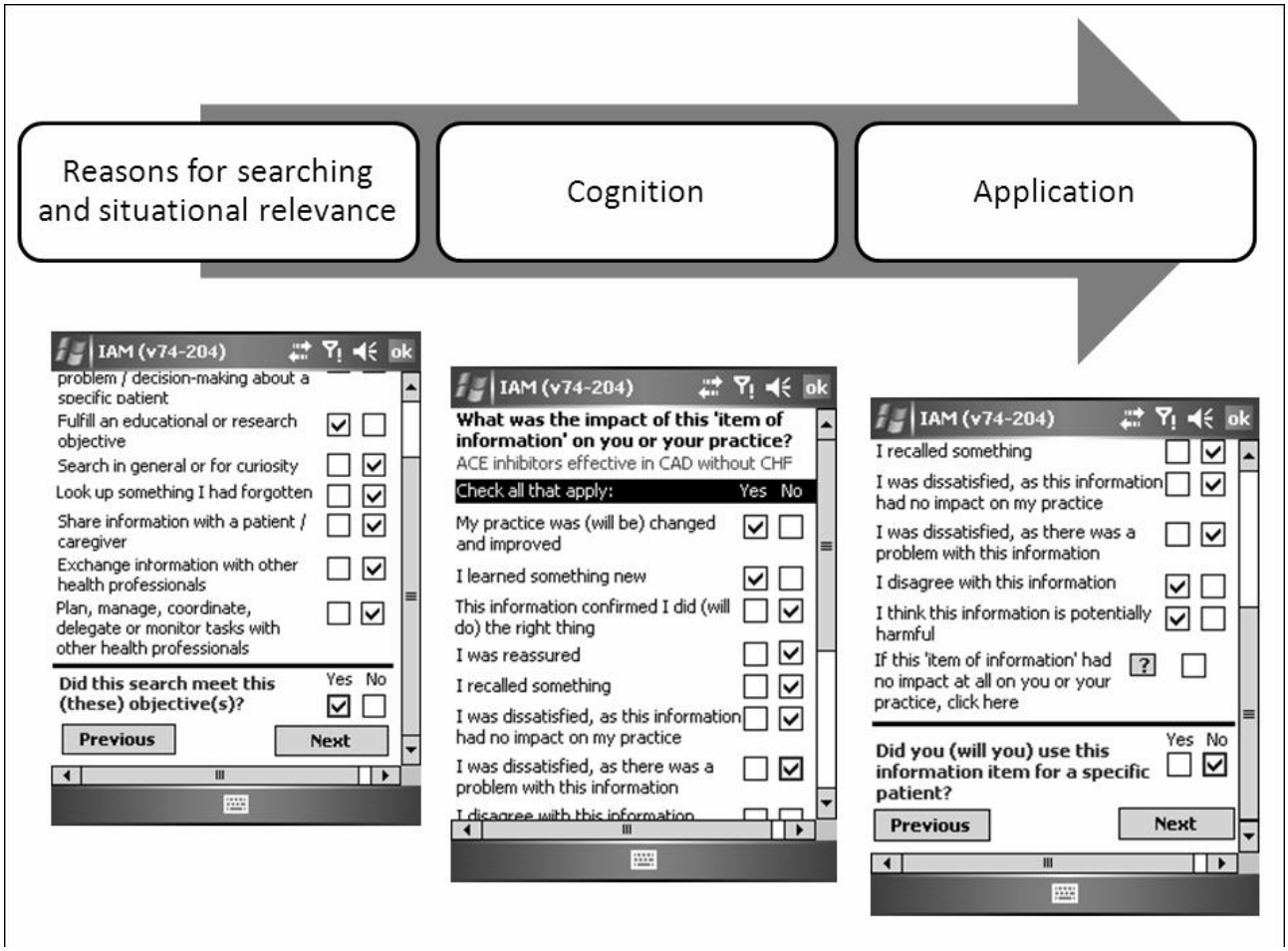
We maintained two separate websites (push and pull). Using a method described elsewhere, we pushed titles of newly released POEMs, hereafter referred to as synopses, to participants by email on weekdays beginning January 7, 2008 [12]. Participants only read synopses they wished to read after clicking on a link in the email message. Ratings of these emailed synopses were also collected at our push website. Ratings were made using the Information Assessment Method (IAM) (described below), and participants earned continuing education credits for this activity, which has been accredited in Canada since 2006. This method, IAM, is a product of our funded research program [22].

XSL•FO
RenderX

To enable and track retrieval of these synopses, each participant received a handheld computer, that is, a personal digital assistant (PDA) or Smartphone containing Essential Evidence Plus. We performed the initial software installation so the device was ready to go on delivery. We specifically chose the PDA for several reasons. First, as a single-user device, a PDA facilitates data collection by attributing information hits to one user. Second, many family physicians are willing to use PDA software for addressing questions arising in their practice. While all participants were offered the HTC Touch Smartphone, 17 chose a PDA with no phone, the hp iPAQ 110. All devices ran the Windows Mobile 6 operating system and were Wi-Fi enabled. However, no data plan was provided and PDA software was used offline.

On each PDA, IAM integrated with Essential Evidence Plus to track all opened information hits as well as the date and time of each search. Using a checklist of seven reasons, IAM prompted each participant to report the reason for their search [23]. IAM then asked the participant to rate the retrieved information in relation to three constructs: (1) situational relevance, (2) cognitive impact, and (3) use of the retrieved information for a specific patient. Figure 2 below shows screen shots from the IAM questionnaire and their corresponding constructs. Participants were trained to use Essential Evidence Plus, and their IAM ratings were transferred to our pull website when their PDA synced with their personal computer (PC). Participants entered the study from November 2007 through May 2008. Each participant had a unique start date defined by the date of their first rated search. Data collection ended in March 2009.

Figure 2. Questionnaire from Information Asessement Method (IAM) linked to one search in Essential Evidence Plus.



Data Analysis

With respect to our first objective, the reading of synopses was tracked in push and pull. Each read synopsis was date and time stamped and attributed to a specific participant. All retrieved synopses previously read as an email alert were classified as dyads, regardless of whether that synopsis was rated. We calculated the number of dyads in total and by participant as well as the time to their occurrence based on the date the synopsis was first read on email.

Qualitative Methods With Respect to Objective 2: Do Family Physicians Purposefully Retrieve Synopses Previously Read on Email

Data Collection

With respect to our second objective, each dyad was a case. On a weekly basis, push and pull databases were merged to identify dyad occurrences. When a dyad was identified, an interview was scheduled and conducted by author JJL. Interviews were recorded on audiotape and transcribed verbatim. Brief semistructured telephone interviews (lasting 16 minutes on

XSL•FO
RenderX

average) were conducted from March 2008 through February 2009. The average time from dyad occurrence to interview was 31 days (range 4-110 days). A longer time to interview was explained by a number of factors, such as delays in synchronizing the handheld computer for data transfer.

Even though we had the dyad concept in mind, interviews were exploratory and began with an open-ended screening question, "Do you remember why you retrieved this POEM?" The purpose of these questions was to identify dyad-related searches that were clearly remembered. This exploratory approach also allowed us to uncover other reasons why a dyad occurred. If the physician's memory of the reason for searching was unclear, the interview ended.

Guided by their personal portfolio of synopsis ratings (quantitative data), which served to remind the participant of the context around the retrieval of specific synopses, interviewees recounted their story around the search. The interview focused on (1) why the search was done, (2) the cognitive impact of information they retrieved, and (3) any application of that information for a specific patient. They were also questioned about perceived patient outcomes. (Our interview guide is available on request.)

Data Analysis

We defined the concept of known items in line with Allen's description in which a user is trying to find an item previously read [24]. Qualitative data consisted of synopses that were read, documents (interviewees' portfolios including ratings and free-text comments on synopses), field notes, and interview transcripts. A thematic analysis was conducted [25]. Text files of transcribed interviews were imported into specialized software (NVivo7, QSR International, Victoria, Australia). Extracts of interviews were assigned by two of the authors (PP and JJL) to emerging themes as suggested by the data. Based on these themes, initially there were three types of dyad: (1) known item, (2) serendipitous, and (3) critical thinking. After group discussion, initial dyad types were refined and organized into two categories (purposeful and serendipitous), each with two subcategories (1a) purposeful, known item, (1b) purposeful, critical thinking, (2a) serendipitous, recognized when reread, and (2b) serendipitous, not recognized when reread. The difference between purposeful and serendipitous information retrieval can be described as follows. In contrast to purposeful retrieval, in a serendipitous encounter the user finds information *not by intention* and the existence or location of information is unexpected [26,27].

For each dyad, five researchers independently assigned interview extracts to dyad types as suggested by the data. These assignments consisted of an iterative process until consensual understanding was achieved. Disagreements were resolved by discussion during consensus meetings.

Results

Quantitative Results With Respect to Objective 1

In the push component, participants had the opportunity to read and rate 194 synopses delivered from January 7, 2008, through December 12, 2008 (or 7814 total opportunities). In this 341-day time window, we documented 4937/7814 (63%) synopsis readings and 4548/7814 (58%) synopsis ratings. On average, 111 synopses were rated per participant (range 11 to 189 ratings). No cognitive impact was reported in 1018 synopsis ratings, while 3530 synopsis ratings contained one or more types of cognitive impact. These ratings are summarized in Table 1.

Table 1. Push: Reports of cognitive impact by type

Type of cognitive impact[a]	n
I learned something new.	2543
I was reassured.	1637
I am motivated to learn more.	1570
This information confirmed I did (will do) the right thing.	1419
This information had no impact at all on me or my practice.	1018
I am reminded of something I already knew.	942
My practice was (will be) changed and improved.	922
I am dissatisfied, as there is a problem with this information.	258
I am dissatisfied, as this information has no impact on my practice.	126
I think this information is potentially harmful.	65
I disagree with this information.	37

[a]More than one type of cognitive impact could be reported for each synopsis.

In the pull component, searches were tracked over a mean of 320 days of follow-up (range 43 to 428 days). We documented 2170 searches in Essential Evidence Plus, and in these searches, 1205 synopses were retrieved. Participants' reasons for searching are reported in Table 2.

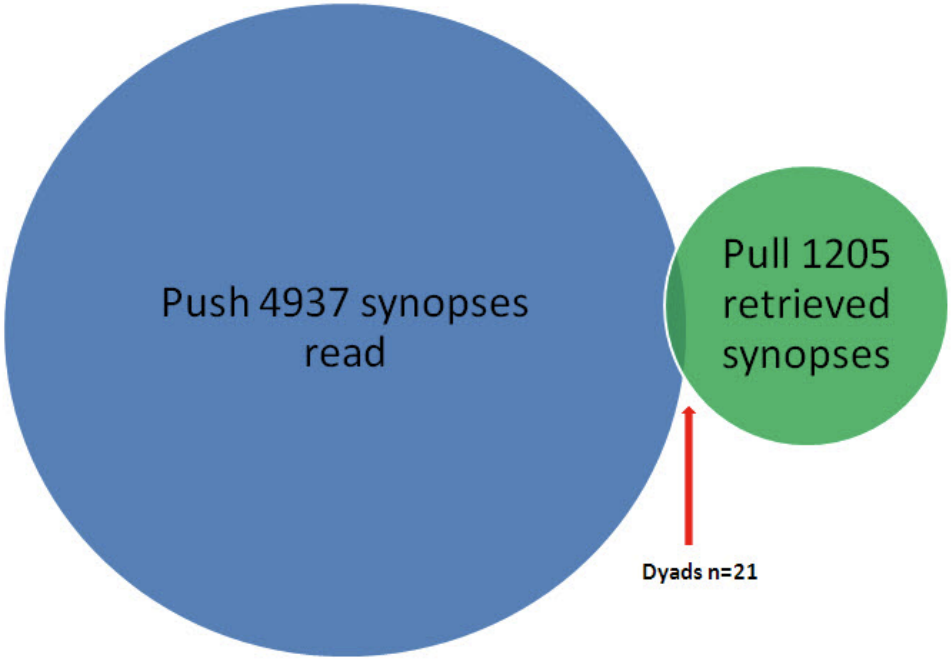
XSL•FO
RenderX

Table 2. Participants' reasons for searches for clinical information

Reason for searching[a]	n
Address a clinical question, problem, or decision about a specific patient	1310
Look up something I had forgotten	672
Share information with a patient/caregiver	624
Exchange information with other health professionals	520
Search in general or for curiosity	496
Fulfill an educational or research objective	434
Plan, manage, coordinate, delegate, or monitor tasks with other health professionals	197

[a]More than one reason could be reported per search.

Figure 3. Number of dyads in the context of synopses read in push and subsequently retrieved. Where dyad signifies one participant retrieving one synopsis he or she previously read on email.



Of the 1205, 21 (1.7%) retrieved synopses were dyads made by 17 participants. Of these, 13 participants made 1 dyad, 3 participants were responsible for 2 dyads each, and 1 participant made 3.

Qualitative Findings With Respect to Objective 2

All 17 participants were interviewed about their dyads, a detailed example of which is presented as Multimedia Appendix 1. Of the 17 participants, one did not clearly remember the dyad.

Purposeful, Known Item Dyads

Of 21 dyads, 6 (28%) were concordant with the known item type of dyad, defined as a search for one synopsis previously read on email. By way of illustration, a participant read and rated a synopsis entitled "Single dose of honey effective for cough in kids" and stated:

> *I wanted to have a copy…for teaching purposes….Well I knew it existed [the synopsis]. When I first read it, I did not write where the article was from to be able to retrieve it. So I had to retrieve it to find which journal it was in*

This extract was interpreted as a known item since the synopsis was retrieved by the participant on purpose because they knew it existed.

Purposeful, Critical Thinking Dyads

Of 21 dyads, 4 (19%) were concordant with the critical thinking type of dyad, defined as a subject search triggered by the content of one pushed synopsis leading to retrieval of other information including that synopsis. In line with Mitchell et al, critical thinking refers to questioning the credibility of clinical information—that is, the accuracy or trustworthiness of clinical information [28]. For example, a participant read one synopsis entitled "Liquid-based equals conventional cervical cytology" and stated:

> *I read that synopsis, and I was very surprised. So I went looking for more information on Pap smears, how accurate they were and more evidence-based*

XSL•FO
RenderX

> *material. I did that through Essential Evidence Plus and through Google. And then I went back to review that synopsis, to make sure I understood what I had read*

This extract suggests the participant was surprised by the content of the pushed synopsis, and this surprise prompted a search. During the search, the participant then retrieved the same synopsis on PDA.

Serendipitous Dyads, Recognized When Reread

Of 21 dyads, 3 (14%) were concordant with our proposed subcategory, serendipitous dyad, recognized when reread. This subcategory is defined as a synopsis retrieved during a subject search on a related topic and clearly recognized when reread. For example, a participant read a synopsis entitled "OCs not associated with overall cancer risks" and stated:

> *I did a search on the oral contraceptive pill.…It was an educational sort of thing I wanted to do for myself....A case came up in the office about birth control....At the time I had forgotten that I had read it [the synopsis]....Then, when I actually read it, I recognized it was something I had read previously*

Serendipitous Dyads, Not Recognized at All

Of 21 dyads, 7 (33%) were concordant with our proposed serendipitous dyad, not recognized at all, defined as a synopsis retrieved in a subject search on a related topic but not clearly recognized when reread or not recognized at all. By way of illustration, a participant read a synopsis entitled "Breastfeeding does not decrease asthma/allergy" and stated:

> *I was actually looking for some information because I did have a patient who asked me about breastfeeding and allergy…. No, I don't remember [having previously read this synopsis on email].*

Findings From Mixing Quantitative and Qualitative Data

Critical thinking dyads are unique in so far as they occurred on the same day that the emailed synopsis was read. No pattern is apparent with respect to the timing of the other dyad types, as shown in Table 3.

Table 3. Number of dyads by type and time of occurrence

Type of Dyad	Time Interval Between Reading in Push and Pull (Days)
Purposeful dyads, known item (n = 6)	0 to 323
Purposeful dyads, critical thinking (n = 3)	Same day
Serendipitous dyads, recognized when reread (n = 3)	22 to 87
Serendipitous dyads, not recognized (n = 5)	19 to 317
Excluded dyad, forgotten (n = 1)	106

Discussion

In clinical medicine, how the push of synopses of clinical research leads to their retrieval is examined in this study. In two situations (known item and critical thinking), family physicians purposefully retrieved a synopsis they had previously read as email. Although the combination of quantitative results and qualitative findings suggests dyads are rare events representing a very small proportion of retrieved information, their occurrence supports our push-pull framework. The rarity of dyads arises from a range of contributing factors outlined in the flow diagram shown in Figure 4.

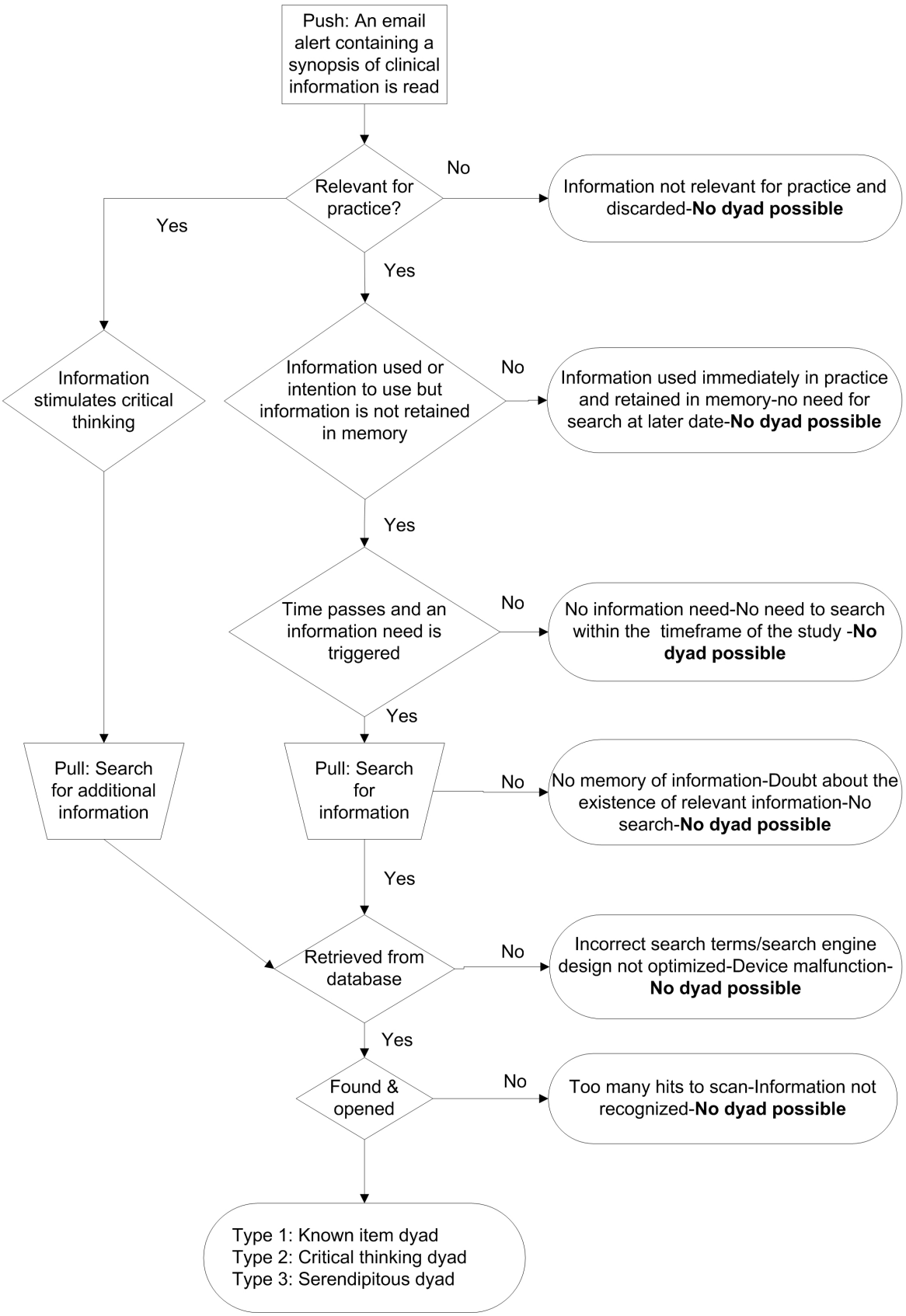
The value of linking the push with the pull of research-based information for practice has been proposed [29-31]. However, our literature review and findings from this study reveal that push and pull are largely treated as separate but important processes. This separation of push and pull can paradoxically complicate the use of clinical information in practice. For example, within a typical primary care patient visit, a known-item search for a synopsis about the dose of metformin for prevention of type 2 diabetes yields such a large set of results that the clinician cannot locate the "needle in the haystack." Our findings suggest a need for a simple method to permit physicians to label a synopsis as a favorite. This would facilitate the creation of user-specific subsets of favorite synopses. Searches for known items within these subsets would be less time consuming and more successful than searching an entire database. In at least one clinical resource, users can presently save synopses delivered as email alerts in a favorites list [32].

Built over years of training and experience, physician memory of clinical information is a critical aspect of any search for known items. In addition to brief reading of clinical information, interactions with colleagues, local opinion leaders, and pharmaceutical representatives are experiences shared by many physicians. In theory, a time-pressured physician needs an efficient search strategy such as known-item searching. However, the capacity to conduct known-item searching is dependent on long-term memory of a specific object of information. Long-term memory can be roughly divided into episodic, semantic, and procedural memory [33]. Semantic memory is the memory of our general knowledge about the world and includes remembering specific information such as facts derived from reading text. When looking for information, semantic memory is called upon for known items of information such as synopses read on email. However, a single exposure to one synopsis on email is a small stimulus to memory, especially as each new day brings the delivery of one new synopsis. Thus, over time, factual knowledge derived from brief reading of email alerts of synopses may be simply forgotten. In related work on cognitive processing and memory, we found the ability of family physicians to remember synopses they previously read declined over time [34]. Future research should strive to

XSL•FO
RenderX

help us better understand if the low level of dyads we observed in this study is related to memory, search skills, limitations of push technology, search engine design [35], or simply low demand for clinical information about problems rarely encountered in primary care practice.

Figure 4. Influence of push on pull: Why dyads are so rare.



XSL•FO
RenderX

Limitations

Our work faced sociotechnical limitations. For example, we could not track failed known-item searches or events where participants retrieved a known synopsis from a database other than Essential Evidence Plus on their PDA. In other work, we interviewed physicians who rated synopses they received on email, similar to the push component of this study. In our other work, we found that of 46 physicians, 8 (17%) said they retrieved synopses as archived email, 3 (7%) said they used a database other than Essential Evidence Plus for retrieving synopses, while 1 (2%) printed synopses for rereading [36]. In the current study, some participants reported technical problems with their PDA, making it likely that searches for synopses were occasionally done at a PC workstation rather than on their PDA. Searches done at a PC workstation could not be tracked [37]. All of these factors reduced our ability to document the occurrence of dyads.

On the other hand, two-thirds of our study cohort were family physicians involved in teaching students or residents. Thus, unlike other studies that excluded academic physicians [38], our data were obtained from a select group who were motivated to read and retrieve synopses for teaching or rhetorical purposes. The motivation to read synopses on email and to search for synopses in one handheld knowledge resource in the context of a research study likely increased the frequency of occurrence of dyads. For some participants, rating a POEM may have enhanced memory of that POEM, and semantic memory is a prerequisite for a known item dyad. For other participants, receiving a PDA may have contributed to a Hawthorne effect that influenced the frequency of their searching.

A strength of our mixed-methods study resides in the integration of qualitative findings and quantitative data to examine the push and pull of research-based synopses. First, the quantitative component allowed us to identify rare dyads in the midst of a large number of information delivery and retrieval events. Prospective identification of dyads through careful tracking allowed us to conduct interviews guided by participants' rating of synopses. Finally, the qualitative component provided some understanding of how participants experienced these events.

Conclusion

In conclusion, email alerts of research-based synopses were rarely retrieved. Our findings help us to better understand the effect of push on pull and to improve the integration of research-based information within electronic resources that are increasingly used by clinicians.

Acknowledgments

We wish to thank Professor Allen Shaughnessy for his critical review of the manuscript. We are grateful to Vinita D'Souza, Yuejing Meng, and Gus Scheidt for their expert assistance. Roland Grad is supported by the McGill University Department of Family Medicine and the Herzl Family Practice Centre in Montreal. Pierre Pluye is supported by a New Investigator Award from the Canadian Institutes of Health Research. This study was funded by a research grant from the Canadian Institutes of Health Research (Knowledge-to-Action reference number KAL-82602).

Conflicts of Interest

None declared

Multimedia Appendix 1

Example of a purposeful known item dyad.

[PDF file (Adobe PDF File), 56 KB - jmir_v13i4e101_app1.pdf]

References

1. Shuval K, Shachak A, Linn S, Brezis M, Reis S. Evaluating primary care doctors' evidence-based medicine skills in a busy clinical setting. J Eval Clin Pract 2007 Aug;13(4):576-580. [doi: 10.1111/j.1365-2753.2007.00805.x] [Medline: 17683299]
2. Glasziou PP. Information overload: what's behind it, what's beyond it? Med J Aust 2008 Jul 21;189(2):84-85 [FREE Full text] [Medline: 18637773]
3. Davidoff F. Reports of reports: how good are secondary publications in medicine? CMAJ 2001 May 29;164(11):1580-1581 [FREE Full text] [Medline: 11402798]
4. Randell R, Mitchell N, Thompson C, McCaughan D, Dowding D. From pull to push: understanding nurses' information needs. Health Informatics J 2009 Jun;15(2):75-85. [doi: 10.1177/1460458209102969] [Medline: 19474221]
5. Strayer SM, Shaughnessy AF, Yew KS, Stephens MB, Slawson DC. Updating clinical knowledge: an evaluation of current information alerting services. Int J Med Inform 2010 Dec;79(12):824-831. [doi: 10.1016/j.ijmedinf.2010.08.004] [Medline: 20951081]
6. Ebell MH. How to find answers to clinical questions. Am Fam Physician 2009 Feb 15;79(4):293-296 [FREE Full text] [Medline: 19235495]
7. Essential Evidence Plus. Browse Databases and Tools. Hoboken, NJ: John Wiley and Sons, Inc URL: http://www.essentialevidenceplus.com/content/poems [accessed 2011-05-02] [WebCite Cache ID 5yO4saP3m]

XSL•FO
RenderX

8. Davies S, Cleave-Hogg D. Continuing medical education should be offered by both e-mail and regular mail: a survey of Ontario anesthesiologists. Can J Anaesth 2004 May;51(5):444-448. [doi: 10.1007/BF03018306] [Medline: 15128629]
9. Johnson ED, Pancoast PE, Mitchell JA, Shyu CR. Design and evaluation of a personal digital assistant- based alerting service for clinicians. J Med Libr Assoc 2004 Oct;92(4):438-444. [Medline: 15494759]
10. Haynes RB, Holland J, Cotoi C, McKinlay RJ, Wilczynski NL, Walters LA, et al. McMaster PLUS: a cluster randomized clinical trial of an intervention to accelerate clinical use of evidence-based information from digital libraries. J Am Med Inform Assoc 2006;13(6):593-600 [FREE Full text] [doi: 10.1197/jamia.M2158] [Medline: 16929034]
11. Mukohara K, Schwartz MD. Electronic delivery of research summaries for academic generalist doctors: a randomised trial of an educational intervention. Med Educ 2005 Apr;39(4):402-409. [doi: 10.1111/j.1365-2929.2005.02109.x] [Medline: 15813763]
12. Grad RM, Pluye P, Mercer J, Marlow B, Beauchamp ME, Shulha M, et al. Impact of research-based synopses delivered as daily e-mail: a prospective observational study. J Am Med Inform Assoc 2008;15(2):240-245 [FREE Full text] [doi: 10.1197/jamia.M2563] [Medline: 18096905]
13. Woodall WG, Buller DB, Saba L, Zimmerman D, Waters E, Hines JM, et al. Effect of emailed messages on return use of a nutrition education website and subsequent changes in dietary behavior. J Med Internet Res 2007;9(3):e27 [FREE Full text] [doi: 10.2196/jmir.9.3.e27] [Medline: 17942389]
14. Fletcher H. Target Marketing. 2009. Subscribing to Hearst's E-mail Philosophy URL: http://www.targetmarketingmag.com/article/new-subscriptions-traffic-drives-hearsts-subject-line-404919/1# [accessed 2011-11-06] [WebCite Cache ID 62zL8JZtu]
15. Magill K. Multichannel Merchant. 2008 May 01. The religion of relevancy URL: http://multichannelmerchant.com/ecommerce/0501-religion-relevancy/ [accessed 2011-11-06] [WebCite Cache ID 62zLhP8BW]
16. Furness V. Global Network. 2008 Apr 01. The Revolution Masterclass on email marketing URL: http://www.marketingmagazine.co.uk/news/798582/Revolution-Masterclass-email-marketing/?DCMP=ILC-SEARCH [accessed 2011-11-06] [WebCite Cache ID 62zM7xcKX]
17. Slone D. Encounters with the library OPAC: Online Searching in Public Libraries. Journal of the American Society for Information Science 2000;51(8):757-773. [doi: 10.1002/(SICI)1097-4571(2000)51:8<757::AID-ASI80>3.0.CO;2-T]
18. Pluye P, Grad RM, Granikov V, Jagosh J, Leung K. Evaluation of email alerts in practice: Part 1. Review of the literature on clinical emailing channels. J Eval Clin Pract 2010 Dec;16(6):1227-1235. [doi: 10.1111/j.1365-2753.2009.001301.x] [Medline: 20722885]
19. Wildemuth BM, O'Neill AL. The "Known" in Known-Item Searches: Empirical Support for User-Centered Design. College and Research Libraries 1995;56(3):265-281.
20. Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. Answering physicians' clinical questions: obstacles and potential solutions. J Am Med Inform Assoc 2005;12(2):217-224 [FREE Full text] [doi: 10.1197/jamia.M1608] [Medline: 15561792]
21. Creswell JW, Plano-Clark VL. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage; 2007.
22. Pluye P, Grad RM, Repchinsky C, Farrell B, Johnson-Lafleur J, Bambrick T, et al. IAM: A Comprehensive and Systematic Information Assessment Method for Electronic Knowledge Resources. In: Dwivedi A, editor. Handbook of Research on Information Technology Management and Clinical Data Administration in Healthcare. Hull, United Kingdom: IGI Global; 2009:521-548.
23. Pluye P, Grad RM, Dawes M, Bartlett JC. Seven reasons why health professionals search clinical information-retrieval technology (CIRT): toward an organizational model. J Eval Clin Pract 2007 Feb;13(1):39-49. [doi: 10.1111/j.1365-2753.2006.00646.x] [Medline: 17286722]
24. Allen B. Recall cues in known-item retrieval. Journal of the American Society for Information Science 1989;40(4):246-252. [doi: 10.1002/(SICI)1097-4571(198907)40:4<246::AID-ASI4>3.0.CO;2-Z]
25. Paillé P. De l'analyse qualitative en général et de l'analyse thématique en particulier. Recherches Qualitatives 1996;15:179-196.
26. Toms EG. Serendipitous Information Retrieval. 2000 Presented at: First DELOS workshop on Information Seeking, Searching and Querying in Digital Libraries; Dec 11-12, 2000; Zurich, Switzerland.
27. Foster A, Ford N. Serendipity and information seeking: an empirical study. Journal of Documentation 2003;59(3):321-340.
28. Mitchell R, Regan-Smith M, Fisher MA, Knox I, Lambert DR. A new measure of the cognitive, metacognitive, and experiential aspects of residents' learning. Acad Med 2009 Jul;84(7):918-926. [doi: 10.1097/ACM.0b013e3181a85609] [Medline: 19550190]
29. Slawson DC, Shaughnessy AF. Teaching evidence-based medicine: should we be teaching information management instead? Acad Med 2005 Jul;80(7):685-689. [Medline: 15980087]
30. Shaughnessy AF, Slawson DC. Are we providing doctors with the training and tools for lifelong learning?. Interview by Abi Berger. BMJ 1999 Nov 13;319(7220):1280 [FREE Full text] [Medline: 10559034]
31. Slawson DC, Shaughnessy AF. Becoming an information master: using POEMs to change practice with confidence. Patient-Oriented Evidence that Matters. J Fam Pract 2000 Jan;49(1):63-67. [Medline: 10678341]
32. BMJ Group. Evidence Updates: From the BMJ Evidence Centre. 2011 Nov 06. URL: http://plus.mcmaster.ca/EvidenceUpdates/ [accessed 2011-11-06] [WebCite Cache ID 631LJLFy2]

XSL•FO
RenderX

33. Atkinson RC, Shiffrin RM. Human memory: A proposed system and its control processes. In: Spence KW, Spence JT, editors. The Psychology of Learning and Motivation: Advances in Research and Theory. Volume 2. New York, NY: Academic Press; 1968:89-195.
34. Grad RM, Pluye P, Granikov V, Liebowitz J, Johnson-Lafleur J, Bartlett G, et al. Does the Information Assessment Method Help to Preserve Long Term Memory of Email Alerts? Family Medicine 2010;42(Suppl 2).
35. Lau AY, Coiera E, Zrimec T, Compton P. Clinician search behaviors may be influenced by search engine design. J Med Internet Res 2010;12(2):e25 [FREE Full text] [doi: 10.2196/jmir.1396] [Medline: 20601351]
36. Pluye P, Grad RM, Johnson-Lafleur J, Bambrick T, Burnand B, Mercer J, et al. Evaluation of email alerts in practice: Part 2. Validation of the information assessment method. J Eval Clin Pract 2010 Dec;16(6):1236-1243. [doi: 10.1111/j.1365-2753.2009.01313.x] [Medline: 20722882]
37. Grad R, Pluye P, Granikov V, Johnson-Lafleur J. Many family physicians will not manually update PDA software: an observational study. Inform Prim Care 2009;17(4):225-230. [Medline: 20359400]
38. Gorman PN, Helfand M. Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered. Med Decis Making 1995;15(2):113-119. [Medline: 7783571]

Abbreviations

CFPC: College of Family Physicians of Canada
IAM: Information Assessment Method
PC: personal computer
PDA: personal digital assistant
POEMS: patient-oriented evidence that matters

Edited by G Eysenbach; submitted 30.11.10; peer-reviewed by H Atherton; comments to author 20.04.11; revised version received 03.05.11; accepted 11.07.11; published 30.11.11.

Please cite as:
Grad R, Pluye P, Johnson-Lafleur J, Granikov V, Shulha M, Bartlett G, Marlow B
Do Family Physicians Retrieve Synopses of Clinical Research Previously Read as Email Alerts?
J Med Internet Res 2011;13(4):e101
URL: http://www.jmir.org/2011/4/e101/
doi:10.2196/jmir.1683
PMID:22130465

©Roland Grad, Pierre Pluye, Janique Johnson-Lafleur, Vera Granikov, Michael Shulha, Gillian Bartlett, Bernard Marlow. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of an eHealth Literacy Intervention for Older Adults

Bo Xie[1], BS, MS, PhD

University of Maryland, College of Information Studies, College Park, MD, United States

Corresponding Author:
Bo Xie, BS, MS, PhD
University of Maryland
College of Information Studies
2117G Hornbake, South Wing
College Park, MD, 20740
United States
Phone: 1 301 405 8617
Fax: 1 301 314 9145
Email: boxie@umd.edu

Abstract

Background: Older adults generally have low health and computer literacies, making it challenging for them to function well in the eHealth era where technology is increasingly being used in health care. Little is known about effective interventions and strategies for improving the eHealth literacy of the older population.

Objective: The objective of this study was to examine the effects of a theory-driven eHealth literacy intervention for older adults.

Methods: The experimental design was a 2 × 2 mixed factorial design with learning method (collaborative; individualistic) as the between-participants variable and time of measurement (pre; post) as the within-participants variable. A total of 146 older adults aged 56–91 (mean 69.99, SD 8.12) participated in this study during February to May 2011. The intervention involved 2 weeks of learning about using the National Institutes of Health's SeniorHealth.gov website to access reliable health information. The intervention took place at public libraries. Participants were randomly assigned to either experimental condition (collaborative: n = 72; individualistic: n = 74).

Results: Overall, participants' knowledge, skills, and eHealth literacy efficacy all improved significantly from pre to post intervention ($P < .001$ in all cases; effect sizes were >0.8 with statistical power of 1.00 even at the .01 level in all cases). When controlling for baseline differences, no significant main effect of the learning method was found on computer/Web knowledge, skills, or eHealth literacy efficacy. Thus, collaborative learning did not differ from individualistic learning in affecting the learning outcomes. No significant interaction effect of learning method and time of measurement was found. Group composition based on gender, familiarity with peers, or prior computer experience had no significant main or interaction effect on the learning outcomes. Regardless of the specific learning method used, participants had overwhelmingly positive attitudes toward the intervention and reported positive changes in participation in their own health care as a result of the intervention.

Conclusions: The findings provide strong evidence that the eHealth literacy intervention tested in this study, regardless of the specific learning method used, significantly improved knowledge, skills, and eHealth literacy efficacy from pre to post intervention, was positively perceived by participants, and led to positive changes in their own health care. Collaborative learning did not differ from individualistic learning in affecting the learning outcomes, suggesting the previously widely reported advantages of collaborative over individualistic learning may not be easily applied to the older population in informal settings, though several confounding factors might have contributed to this finding (ie, the largely inexperienced computer user composition of the study sample, potential instructor effect, and ceiling effect). Further research is necessary before a more firm conclusion can be drawn. These findings contribute to the literatures on adult learning, social interdependence theory, and health literacy.

(J Med Internet Res 2011;13(4):e90) doi:10.2196/jmir.1880

KEYWORDS

Health literacy; lifelong learning; aged; technology

XSL•FO
RenderX

Introduction

Health literacy is "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" [1]. This concept has drawn much attention recently [2-6], with increasing well-documented evidence of both the negative impact of poor health literacy on health outcomes and health care costs [7] and the alarmingly low levels of health literacy among American adults: a national survey showed only 12% of the adults in the United States have proficient health literacy, and this proportion drops to 3% among older adults [8].

Recently, information and communication technologies (ICTs) are increasingly being widely used in health care [9,10], presenting both opportunities and challenges for developing and implementing effective health literacy interventions. As government agencies such as the US National Institutes of Health (NIH), nonprofit organizations such as medical associations, and for-profit organizations alike are increasingly putting health information online, the Internet has already become an invaluable resource for high-quality health information [11-13]. This resource, however, can only be useful if the user has adequate eHealth literacy, or "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [14]. Individuals who have low health literacy—for instance, older adults—are likely to also have low computer and Internet literacy [15-17], thus facing a double jeopardy in the eHealth era.

Existing literature provides little scientific evidence about effective health literacy interventions [18], and even less about effective interventions for improving the eHealth literacy of the older population. Existing interventions focus predominantly on simplifying medical materials and instructions [19-21]. This "lowering-the-bar" approach is useful but limited given the complexity of medical terminology and knowledge. Education and training is another key approach to addressing the health illiteracy crisis [22]. This approach requires an understanding of health literacy as an active, lifelong learning process that features continuous learning of new, valid health information and unlearning of outdated, harmful information [22]. Such an understanding is especially important in the context of ICTs being increasingly used by health consumers, professionals, and policy makers alike in health care [9,10]. As ICTs change at a rapid rate, so do the requirements for health literacy skills [23].

Electronic Health Information for Lifelong Learners

The present study is a part of the Electronic Health Information for Lifelong Learners (*e*HiLL) research project that aims to address these gaps in the literature [4-6,24]. The goal of the larger *e*HiLL research project is to generate scientific knowledge about optimal learning conditions and strategies that can effectively and efficiently improve older adults' learning and use of eHealth applications. To achieve this goal, the *e*HiLL research project consists of a series of experimental studies designed to examine the effects of various learning conditions and strategies through theory-driven, hypothesis-testing, rigorous experiments. The *e*HiLL experimental studies build on an understanding of health literacy as an active, lifelong learning process that goes beyond the formal educational settings in early life stages [22]. Importantly, guided by the literature on older adults' learning of computer technology, the *e*HiLL interventions all feature key elements designed specifically to accommodate older computer learners' needs and preferences [4]. These include (1) providing step-by-step, detailed instructions and avoiding technical jargon [24,25], (2) providing hands-on practice and encouraging questions [26], (3) making sure each lesson builds on previous lessons and increases complexity gradually [25,27], (4) ensuring the learners experience at least some level of success at the initial stage of the training [25,28,29], (5) conducting the training in a familiar, relaxed, and supportive environment [27,29], and (6) offering the training in the early morning hours, which is generally the optimal time of day for older learners [30].

These key elements were fully incorporated in prior *e*HiLL interventions and proven to be effective in improving older adults' eHealth literacy [4-6,24]. Building on and expanding the success of these prior studies, the present study fully incorporates these key elements while adding a new aspect—that is, collaborative versus individualistic learning. Existing literatures on adult learning and cognitive development in later life provide a theoretical foundation for the collaborative versus individualistic versions of the intervention tested in the present study. The present study differs from prior *e*HiLL studies in important ways. First, the only prior *e*HiLL study [6] that involved both collaborative and individualistic learning had a brief intervention time (the experimental session lasted 2 hours) and it used a 16-minute-long video tutorial as the curriculum. In comparison, the present study involved 2 weeks of intervention for a total of 8 hours, using a paper-based curriculum. Second, while the present study and prior *e*HiLL studies [4,5,24] all used instructional materials drawn from the same tutorial, the Xie and Bugg [4] and Xie [24] studies focused on individualistic learning and the Xie [5] study was only on collaborative learning. The present study is the first *e*HiLL study that compares the relative effects of collaborative versus individualistic learning over an extended period of intervention time.

Collaborative Versus Individualistic Learning

Collaborative learning is one of the most common forms of active learning. It can be defined as "any instructional method in which students work together in small groups toward a common goal," ([31] p 223). Collaborative learning requires learners to actively engage in the learning process by engaging in meaningful activities and reflecting on what they are learning from doing those activities [32]. Collaborative learning is often contrasted with individualistic learning that features students working on their own with little or no interaction with peer students [33].

The superiority of collaborative learning over individualistic learning is predicted by social interdependence theory, which emphasizes the interdependence among group members by arguing that the group is a "dynamic whole," such that any

XSL•FO
RenderX

change in the state of a group member changes that of other group members [33]. The social interdependence among group members can be positive, negative, or nonexistent. Positive interdependence (collaboration) facilitates learning by promoting collaboration among group members. It can be found when individuals recognize that the only way they can achieve their goals is when other group member also achieve their goals. Negative interdependence (competition) exists when individuals recognize that the only way they can achieve their goals is when others fail to do so. It often results in obstructive interactions impeding learning. Nonexistent interdependence (individualistic efforts) exists when individuals perceive that their achievement is not affected by others' performance. It features no interaction among group members as each member learns independently [33,34].

Extensive empirical evidence supports the effectiveness of collaborative learning. A meta-analysis of over 300 studies shows that collaborative learning outperforms individualistic and competitive learning in postsecondary and professional settings [34]. However, there are still major gaps in the literature that require further examination. Collaborative learning research within the social interdependence tradition is predominantly based on formal education of younger adults [33,34]. Whether these findings can be generalized to older age groups in informal educational settings is yet to be answered.

Recently, the cognitive-developmental literature has begun to examine collaborative learning as a mechanism for improving cognitive abilities in later life [35-37]. Some (non-computer-related) studies find a positive impact of collaborative learning on older adults' performance [38,39]. However, there is also evidence that, compared with individualistic learning, collaborative learning has no, or even negative, impact [40-42]. A possible reason might be that existing research within the cognitive-developmental tradition generally does not provide detailed instructions to ensure collaboration [42]. Instead, participants are simply instructed to "work together" [43] or "collaborate as much as possible" [40]. To ensure collaborative learning for older adults in informal settings, it is critical to develop effective strategies that really work.

Group Composition

Group composition may affect the "dynamic whole" of a group and, subsequently, the learning process and outcomes [33,34]. Evidence suggests there is more collaboration in groups with either female- or male-gender majority than in groups with equal gender composition [44,45], and more collaboration in same-gender groups than in mixed ones [44,46].

The time and effort spent on getting familiar with each other and coordinating may negatively affect the learning process and outcomes [43]. This argument finds support in research reporting collaborative learning with familiar partners (typically defined as related family members such as a spouse) being more effective than that with unfamiliar partners in enhancing cognitive performance [47] or in reducing the negative effects of collaboration [40,41]. Yet evidence exists that, even with familiar partners, collaborative learning does not generate more benefits than individualistic learning in improving older adults' cognitive performance [42].

Ample evidence suggests that prior computer experience is a strong predictor for older adults' computer task performance and learning outcomes [48-50]. Cody et al [28] in their study of older computer learners found that "the same [computer training] program was too challenging to some and insufficiently stimulating to others" (p 282). Some researchers [25] suggest forming homogeneous groups based on prior computer experience to ensure the success of computer training for older adults. This suggestion should be taken with caution, given that to date research on computer training for older adults has focused predominantly on individualistic learning while paying little attention to collaborative learning (an important exception is the Zandri and Charness study [51], which found promising signs for the superiority of collaborative learning over individualistic learning). While homogeneous groups based on prior computer experience might work better than heterogeneous groups in individualistic learning, this might not be the case for the collaborative learning condition.

Research Questions and Hypotheses

The present study asked the following primary research question: what impact might the intervention have on older adults' computer and Web knowledge, procedural skills, eHealth literacy efficacy, attitudes, and participation in their own health care?

Under this primary research question, 2 subresearch questions were asked: (1) what impact might the learning method (collaborative; individualistic) have on the learning outcomes?, and (2) what impact might group composition (based on gender, prior familiarity with peers, and prior computer experience) have on the learning outcomes?

The following hypotheses were tested: (1) hypothesis 1: computer/Web knowledge, skills, and eHealth literacy efficacy increase significantly from pre to post intervention (at the .05 level; same for all hypotheses), (2) hypothesis 2: collaborative learning is significantly more effective than individualistic learning in improving the learning outcomes, and (3) hypothesis 3: collaborative learning is significantly more effective than individualistic learning in heterogeneous group compositions, while individualistic learning is more effective than collaborative learning in homogeneous group compositions.

Methods

Design

We used a 2 × 2 mixed factorial design with learning method (collaborative; individualistic) as the between-participants variable and time of measurement (pre; post) as the within-participants variable.

Research Sites

The Hyattsville and New Carrollton branch libraries of the Prince George's County Memorial Library System served as the primary research sites for this study. The Library System is a publicly funded large, urban library system serving over 830,000 residents in Prince George's County, Maryland, USA.

XSL•FO
RenderX

It has been serving as the primary research site for the larger *e*HiLL research project since 2007 [4-6,24]. This Library System was selected as the key site for the *e*HiLL research project because it serves a large population of ethnic minorities, particularly African American/black people. According to the US Census Bureau, 66% of Prince George's County residents are African American/black, much higher than the 30% overall rate of African American/black residents in Maryland or the 12% rate nationwide (http://www.census.gov). Partnering with this Library System ensures the reach of *e*HiLL interventions to individuals from this underserved minority group. The Hyattsville and New Carrollton branch libraries of the Library System provided free networked computers, space, and staff support to facilitate the implementation of this study. The geographic location of these branch libraries is convenient for potential research participants and the researchers. Both are within 10 miles of the University of Maryland and easily accessible by car or public transportation.

Participants

Standard recruitment techniques were used to recruit participants. These included posting recruitment flyers in the branch libraries and other local organizations (eg, senior centers, community centers, and churches) and advertising in the Library System's newsletter. The inclusion criterion was age 60 years and above, though on request and in cases where seats were available, we accommodated individuals a few years younger. A total of 146 older adults aged 56–91 years (mean 69.99, SD 8.12) participated in the present study during a 4-month period (February to May 2011). We randomly assigned 72 participants to the collaborative learning experimental condition and 74 to the individualistic learning condition. Table 1 summarizes participants' basic characteristics, including demographics, health status, and prior computer experience.

XSL•FO
RenderX

Table 1. Participants' basic characteristics

Variable	n	%
Gender		
Female	96	69
Male	44	31
Highest level of education		
Less than high school graduate	15	11
High school graduate/GED[a]	39	28
Vocational training	11	8
Some college/associate's degree	33	23
Bachelor's degree	25	18
Master's degree or other postgraduate training	16	11
Doctoral degree	2	1
Ethnic group		
Asian	8	6
African American/black	90	64
White	30	21
Other	13	4
Household annual income range (US $)		
<20,000	38	28
20,000–29,999	28	20
30,000–39,999	14	10
40,000–49,999	8	6
50,000–59,999	9	6
60,000–69,999	6	4
70,000–99,999	4	3
≥100,000	3	2
Do not know for certain	12	9
Do not wish to answer	18	13
Health status		
Poor	7	5
Fair	24	17
Good	75	53
Very good	25	18
Excellent	11	8
English as primary language		
Yes	125	88
No	16	12
Frequency of computer use		
Every day	12	9
Every 2–3 days	27	20
Once a week	18	13
More than once a month	13	9

XSL•FO
RenderX

Variable	n	%
Less than once a month	20	15
Never	48	35

[a] General equivalency diploma.

Measures

Adapted from existing outcome measures of collaborative learning (that focus on younger learners in formal educational settings) and with necessary modifications based on the results of a prior *e*HiLL study [5], outcome measures (dependent variables) for the present study covered the following categories: knowledge gains, skill gains, efficacy, attitudes, and changes in participation in own health care. Learning preference, familiarity with peers, prior computer experience, and basic demographics were measured to serve as control variables. (Copies of the instrument are available on request to the author.)

Computer/Web Knowledge

This was measured by objective tests of knowledge about components of the computer and the Web. Participants were shown an image of a computer and a screenshot of the NIHSeniorHealth.gov website, and were instructed to write down names of the main components of each image (eg, keyboard, mouse, link, scroll bar). Computer knowledge and Web knowledge were each measured by 5 items; each item scored 1 point if answered correctly and 0 points if answered incorrectly with a scoring range of 0–5.

Computer/Web Skills

These were measured by procedural tests of the abilities to carry out specific computer and Web operations. Each participant had one computer to use during the testing. Participants were asked to perform a number of operations on their computers independently. Participants had up to 1 minute to complete each operation. Sample operations were to open a Web browser; go to the NIHSeniorHealth.gov Web site; increase text size; find information on the Falls and Older Adults health topic; and open a video. There were a total of 20 operations. Each operation scored 1 point if successfully completed and 0 points if unsuccessful, with a scoring range of 0–20.

eHealth Literacy Efficacy

This construct was measured by the eHealth literacy scale [52], which was built on the self-efficacy concept [53] to measure perceived skills at and comfort with using the Internet for health information and decision making. The main scale has 8 items. Each item is on a 1- to 5-point Likert scale with the following anchors: 1: strongly disagree; 2: disagree; 3: undecided; 4: agree; 5: strongly agree. Higher score indicates higher eHealth literacy efficacy. A sample item is "I know **how** to find helpful health resources on the Internet" (bold original). This scale has been used in multicultural samples and has shown excellent internal consistency reliability (scale alpha = .89–.97) with good test–retest reliability [23].

Attitude Toward the Class

This was measured by the following 5 items (items 3 and 4 were modified from Pace and Kuh [54]):

1. Overall, what would you say about the instructor's teaching? (Anchors: 1: very poor; 2: poor; 3: fair; 4: good; 5: excellent)

2. Overall, was this computer class useful to you? (1: completely useless; 2: useless; 3: somewhat useful; 4: useful; 5: very useful)

3. How would you evaluate your entire experience in this computer class? (1: extremely dissatisfied; 2: dissatisfied; 3: neither satisfied nor dissatisfied; 4: satisfied; 5: extremely satisfied)

4. If you could start over again, would you go to the same computer class you are now attending? (1: definitely not; 2: probably not; 3: not sure; 4: yes; 5: definitely yes)

5. Would you want to recommend this training class to other people similar to your age? (1: definitely not; 2: probably not; 3: not sure; 4: yes; 5: definitely yes)

Changes in Participation in Own Health Care

This was measured by 12 items, including 6 items modified from a Kaiser survey study [16], 5 items modified from a Pew survey study [55], and an additional item added to supplement the Kaiser and Pew items. (These items are detailed in Table 6 in the Results section where the results are reported.)

Attitude Toward the Individualistic Versus Collaborative Learning Method

This was measured by the following item: "When I have to learn a new skill, I prefer to learn alone, rather than with others" [51]. It was scored on a 1- to 5-point Likert scale with the following anchors: 1: disagree strongly; 2: disagree; 3: undecided; 4: agree; and 5: agree strongly.

Prior Experience

Prior experience (familiarity) with peers was measured by the following item: "Are you related to or familiar with at least one person taking this same computer class? (eg, spouse, sibling, friend, acquaintance)". Prior computer experience was measured by the frequency of computer use.

Basic Demographics

Age, gender, education, health status, race/ethnicity, income, and primary language were recorded.

Table 2 summarizes these measures and time(s) of measurement.

XSL•FO
RenderX

Table 2. Measures used in the present study and the time(s) of measurement

Variable	Pre	Post
Computer/Web knowledge	X	X
Computer/Web skills	X	X
eHealth literacy efficacy	X	X
Changes in health behavior/decision making		X
Attitude toward the class		X
Attitude toward the individualistic versus collaborative learning method	X	
Prior experience with peers	X	
Prior computer experience	X	
Basic demographics	X	

Instructional Materials

This study used a set of instructional materials developed by the National Institute on Aging (NIA) of the NIH, "Helping Older Adults Search for Health Information Online: A Toolkit for Trainers" [56]. This freely available toolkit is designed to improve older adults' ability to access and use NIH online health information resources (eg, the NIHSeniorHealth.gov website). By focusing on only NIH resources, this *e*HiLL intervention avoids potential problems associated with the quality of online health information [57,58]. The toolkit features detailed lesson plans, in-class interactive exercises, take-home practice exercises, and other supportive handouts (eg, a glossary of computer terms). This toolkit was chosen because, first, it contains key elements that, as addressed in the Introduction section above, facilitate older adults' learning of computer technology [24-30]; and second, it was tested in prior *e*HiLL studies and proven to be effective [4,5,24].

The toolkit contains lesson plans (modules) designed to be used independently or in any combination. We used 4 modules in the present study to help older adults learn about (1) basic computer terms (1 module), (2) NIHSeniorHealth.gov (2 modules), a website designed to accommodate age-related changes in cognitive, physical, and sensory abilities [59], and (3) evaluating the quality of online health information (1 module). Together, these 4 selected modules provide a good coverage of the eHealth literacy skills as defined by Norman and Skinner [14]. Table 3 [56] outlines the lesson plans and goals of these 4 modules.

XSL•FO
RenderX

Table 3. Lesson plans and goals included in the National Institute on Aging (NIA) toolkit (extracted from the toolkit [56]).

Class session	Lesson goals
Session 1: Internet Basics (NIA Module #1)	1. Learn basic computer terms
	2. Practice using the mouse
	3. Learn basic Internet terms
	4. Learn how to get to a website
	5. Learn how to explore a website
	6. Learn how to use a search box
	7. Learn how to use a site map
Session 2: Introduction to NIHSeniorHealth (NIA Module #2)	1. Use the Home Page to find health topics on NIHSeniorHealth
	2. Use the Table of Contents of a health topic to find specific information
	3. Navigate through a health topic
	4. Enlarge, view, and close images
	5. Find answers to health questions of personal interest
Session 3: NIHSeniorHealth Quizzes and Videos (NIA Module #3)	1. Recall how to use the Home Page of the NIHSeniorHealth website
	2. Recall how to use the All Topics A–Z page to find health topics on the NIHSeniorHealth website
	3. Recall how to use the special features (optional)
	4. Learn how to take online quizzes
	5. Learn how to open, watch, and close a video
	6. Learn how to open, read, and close a video transcript
	7. Learn how to find answers to health questions of personal interest
Session 4: Evaluating Health Websites (NIA Module #9)	1. Reliable health information websites
	2. The sponsor of a health website
	3. The purpose of a health website
	4. The authors of the health information
	5. The reviewers of the heath information
	6. The most recent update of the health information
	7. The privacy policy of a health website
	8. Clues about the accuracy of a website's health information
	9. The contact information for a health website

Procedure

The general procedures of the present study are similar to those of prior *e*HiLL studies [4-6,24]. In the first session, participants first signed the consent form (approved by the Institutional Review Board of the University of Maryland). The pre intervention survey questionnaire was then administered, followed by the pre skill testing. The intervention began with the completion of the pre testing. At the end of the last session, the post intervention survey questionnaire was administered, followed by the post skill testing. Each class met twice a week, 2 hours each time between 9:00 and 11:00 am, for a total of 2 weeks, at a library site. Class size was small (no more than 8 participants per class). The instructors, trained Master of Library Science students, frequently provided immediate, positive, and useful feedback when needed. Each participant had one computer to work on during each session. The instructors emphasized hands-on practice and provided relevant handouts during each session. These procedural components were carefully designed based on the literature on older adults' computer learning [24-30] and proven to be effective in prior *e*HiLL studies [4-6,24].

Compared with prior *e*HiLL studies [4-6,24], a unique procedural aspect of the present study was the use of both collaborative and individualistic learning methods. Building on prior work on computer class structure for older learners that focused primarily on individualistic learning [26] and the common strategies used to promote collaborative learning among younger adults in formal learning settings [33,34], and adapting those strategies to accommodate the special needs and preferences of older adult learners in the public library setting, we used several strategies during the sessions to promote individualistic or collaborative learning (Table 4).

XSL•FO
RenderX

Table 4. Class structure and strategies for the learning conditions

Activity/time	Individualistic learning	Collaborative learning
Housekeeping: 5 minutes	• Welcome • Instructor self-introduction • Participants self-introduction • Practical information • How long the class session will last • Where the restrooms are • Environment check • Everyone has a computer • Everyone can see instructor • Everyone can hear	• Same as the individualistic condition
Overview: 5 minutes	• Goal statement • What the participants will know or be able to do after this class session • Agenda • What will happen during this session • Steps and procedure • What instructor will do and what participants will do	• Same as the individualistic condition
Explanation of learning method: 1 minute	• Explain explicitly that everyone in the class is expected to learn independently • Encourage participants to ask the instructor any questions that they might have	• Explain explicitly that the class is expected to learn together as a group • Encourage participants to share with and help peers with any questions that they might have (and explain that instructor will answer any remaining question)
Introduction to the specific topic of this class session: 5 minutes	• Definitions, scope, background information	• Same as the individualistic condition
Lecture and demonstration, step-by-step instruction (part 1): 20 minutes	• Present material and demonstrate processes, following the instructions and examples used in the National Institute on Aging toolkit • Encourage questions • Get confirmation after each step is explained and demonstrated • Check frequently to make sure everyone is on the same page	• Same as the individualistic condition
Brief reflection: 2 minutes	• Pause briefly and instruct participants specifically to check their own notes and reflect independently	• Pause briefly and instruct each participant specifically to compare notes with a peer sitting next to him or her and reflect together with peer

XSL•FO
RenderX

Activity/time	Individualistic learning	Collaborative learning
Continuation of lecture and demonstration: 20 minutes	• Same as part 1 above	• Same as the individualistic condition
Brief reflection: 2 minutes	• Same as the first independent reflection session above	• Same as the first collaborative reflection session above
Break: 5 minutes	• Distribute handouts, which have in-class practice exercises and detailed, step-by-step instructions for completing the exercise	• Same as the individualistic condition
Hands-on practice: 40 minutes	• Participants perform the hands-on practice activity independently	• Each participant pairs up with a peer to do the hands-on practice activity together
	• Encourage participants to ask instructor questions about the specific steps of the exercise	• Encourage participant to ask peers questions about the specific steps of the exercise
	• Help to answer each participant's questions	• Participants discuss with peers and try to answer peers' questions (if, after peer discussion and exploration, questions remain, then instructor will answer the remaining questions)
Practice vs group reflection: 10 minutes	• Same as above (participants continue to engage in hands-on practice independently)	• All participants (and instructor) sit in a circle to discuss, share, and reflect together
Closing: 5 minutes	• Summarize content covered in this class session	• Same as the individualistic condition
	• Distribute handouts for take-home exercises, which have detailed, step-by-step instructions for completing the exercises	
	• Point out opportunities for coming back to use the library's computers to practice	
	• Preview the topic of next class session	
	• Thank participants for coming to this class session and remind them to come to the next class session	

Data Analysis

The main statistical analyses conducted to test the hypotheses were various techniques of analysis of variance (ANOVA). These included multivariate repeated measures analyses, one-way ANOVA, and univariate analysis of covariance. Two-independent-samples tests (Mann-Whitney U) and chi-square tests were used to compare the collaborative versus individualistic samples.

Results

Comparing the Collaborative Versus Individualistic Experimental Groups

Two-independent-samples tests (Mann-Whitney U) found no significant difference in age (P = .13), education (P = .11), health (P = .85), income (P = .32), and computer use frequency (P = .06), and chi-square tests found no significant difference in gender (P = .66), ethnicity (P = .07), primary language (P = .81), and familiarity with peers (P = 1.00) among participants in the collaborative and individualistic experimental conditions. Chi-square test also found no significant difference in the retention rate of the 2 experimental conditions (P = .56); overall, a total of 108 participants completed both the pre and post testing, resulting in a 74% retention rate for this 2-week intervention. These results suggest the 2 experimental groups were comparable in these aspects.

Comparing Participants Who Completed the Intervention Versus Those Who Did Not

Mann-Whitney U tests found no significant difference in age (P = .51), education (P = .41), health (P = .42), income (P = .78), and baseline computer knowledge (P = .80), Web knowledge (P = .81), skills (P = .70), and eHealth literacy efficacy (P = .12) between participants who completed both the pre and post testing and those who completed only the pre testing. Significant difference was found in computer use frequency, with participants who completed both the pre and

XSL•FO
RenderX

post testing reported having *less* prior use of computers than those who completed only the pre testing ($P = .007$).

Changes From Pre to Post Intervention

Multivariate repeated measures analyses found that, overall, computer knowledge, Web knowledge, procedural skills, and eHealth literacy efficacy improved significantly from pre to post intervention ($F_{4,90} = 119.60$, $P < .001$). Univariate repeated measures analyses revealed significant improvement from pre to post intervention on each of these 4 measures ($P < .001$ in all 4 cases; computer knowledge: $F_{1,93} = 60.60$; Web knowledge: $F_{1,93} = 54.92$; procedural skills: $F_{1,93} = 264.40$; and eHealth literacy efficacy: $F_{1,93} = 229.31$). Hypothesis 1 was strongly supported. Further, effect sizes (measured by Cohen's d) with regard to gains from pre to post intervention in computer and Web knowledge, skills, and eHealth literacy efficacy ranged from 0.88 to 2.25. The statistical power of these measures reached 1.00 even at the alpha = .01 level (Table 5).

Table 5. Effect sizes and statistical power

Variable	Cohen's d	Percentile standing	Statistical power (alpha = .01)
Computer knowledge	1.05	84	1.00
Web knowledge	0.88	80	1.00
Procedural skill	1.70	95	1.00
eHealth literacy efficacy	2.25	99	1.00

Attitudes

Participants in both the collaborative and individualistic learning conditions had overwhelmingly positive attitudes toward the intervention. Across the 2 conditions, 99% (94/95) of participants felt the instructor's teaching was good or excellent; 95% (90/95) of participants felt the intervention was useful or very useful; 98% (92/94) were satisfied or extremely satisfied with the entire experience participating in the intervention; 84% (79/94) said they would attend the same class if it started over; and 98% (93/95) would recommend the intervention to their age peers. Multivariate ANOVA found no significant difference between participants in the collaborative and individualistic conditions in all measures of attitude.

It is worth noting, though, that 1 attitudinal measure, the instructor's teaching, *approached* a statistically significant difference between the 2 experimental conditions ($F_{1,104} = 3.34$; $P = .07$). While participants in both conditions had very positive assessment of the instructor's teaching, participants in the individualistic learning condition had a slightly more positive attitude toward the instructor's teaching than those in the collaborative learning condition (individualistic: mean 4.87, SD 0.35; collaborative: mean 4.72, SD 0.45).

Changes in Participation in Own Health Care

Across the collaborative and individualistic learning conditions, a notable number of participants reported changes in various aspects of participation in their own health care (Table 6).

XSL•FO
RenderX

Table 6. Changes in participation in own health care as a result of the intervention

	Yes		No	
	n	%	n	%
Since you started taking this computer class, have you had a conversation with a friend or family member about the health information you found on NIHSeniorHealth?	55	59	39	42
Since you started taking this computer class, have you talked with a doctor or other health care provider about the information you found on NIHSeniorHealth?	9	10	84	90
Have you changed your behavior because of the health information you found on NIHSeniorHealth?	51	55	41	45
Have you made a decision about how to treat an illness or condition because of the information you found on NIHSeniorHealth?	56	61	36	39
Have you changed your health insurance plan because of the information you found on NIHSeniorHealth?	2	2	86	98
Have you changed your overall approach to maintaining your health or the health of someone you help take care of because of the information you found on NIHSeniorHealth?	58	64	33	36
Has the information you learned from NIHSeniorHealth led you to ask a doctor new questions or to get a second opinion from another doctor?	54	62	33	38
Has the information you found on NIHSeniorHealth changed the way you think about diet, exercise, or stress management?	69	76	22	24
Has the information you found on NIHSeniorHealth changed the way you cope with a chronic condition or manage pain?	51	57	38	43
Has the information you found on NIHSeniorHealth affected a decision about whether to see a doctor?	37	42	52	58
Have you changed the way you take medicine because of the information you found on NIHSeniorHealth?	32	36	58	64

One-way ANOVA found no significant difference in the total number of reported changes in participation in own health care between participants in the collaborative and individualistic learning conditions (P = .45).

Collaborative Versus Individualistic Learning Method

Multivariate repeated measures analyses found a significant main effect of the learning method on computer knowledge, Web knowledge, procedural skills, and eHealth literacy ($F_{4,90}$ = 4.56, P = .002). To examine on which specific outcome measure(s) the learning method showed a main effect, univariate repeated measures analyses were performed and revealed a significant main effect of the learning method on procedural skills ($F_{1,93}$ = 7.81; P = .006) and eHealth literacy ($F_{1,93}$ = 8.64; P = .004). Univariate analyses revealed no significant main effect of the learning method on either computer knowledge (P = .51) or Web knowledge (P = .47).

Interestingly, one-way ANOVA found a significant difference in pre intervention procedural skills ($F_{1,142}$ = 7.17; P = .008) and eHealth literacy ($F_{1,140}$ = 6.18; P = .01) between participants in the individualistic and collaborative learning groups. Pre intervention computer knowledge (P = .90) and Web knowledge (P = . 94) did not differ significantly between the 2 experimental groups. To examine whether the significant main effects of the learning method on procedural skills and eHealth literacy as revealed by the univariate analyses were due to significant differences in these variables at the baseline, univariate analysis of covariance was performed and, after controlling for pre intervention procedural skills and eHealth literacy, respectively, the significant effects of the learning method on these variables both disappeared (procedural skills: P = .36; eHealth literacy: P = .06). These findings suggest that, when controlling for these variables at the baseline, the learning method had no main effect on knowledge, skills, and eHealth literacy efficacy. Hypothesis 2 was not supported.

Multivariate repeated measures analyses found no significant interaction effect of the learning method (collaborative; individualistic) and time of measurement (pre; post): P = .73.

Group Compositions

Groups in this study had 4 gender compositions (there was no male-only group): female-only (14 participants in this group composition), female majority (n = 96), equal number of female and male (n = 10), and male majority (n = 13). Multivariate repeated measures analyses found no significant 3- or 2-way interaction effect of group gender composition, learning method, and time of measurement (learning condition × gender composition: P = .40; time × gender composition: P = .74; time × learning condition × gender composition: P = .79). Further, group gender composition had no significant main effect on the outcome measures (P = .68).

Groups had 2 compositions based on familiarity with peers: familiar with at least one other person in the same session (n = 37) and not familiar with anyone in the same session (n = 106). Multivariate repeated measures analyses found no significant 3- or 2-way interaction effect of group familiarity composition, learning method, and time of measurement (learning condition × familiarity composition: P = .37; time × familiarity

XSL•FO
RenderX

composition: P = .80; time × learning condition × familiarity composition: P = .88). Further, no significant main effect of group familiarity composition was found (P = .53).

Similar to 2 prior *e*HiLL studies [5,6], computer use frequency was used to categorize group composition based on prior computer experience. "Experienced computer users" were defined as individuals who use the computer every day or every 2–3 days a week, and "inexperienced computer users" were defined as individuals who use the computer less than every 2–3 days a week. Using this criterion, we coded each session into 1 of the following 5 groups: inexperienced user only (n = 36), inexperienced user majority (n = 88), equal number of experienced and inexperienced users (n = 4), experienced user majority (n = 10), and experienced user only (n = 1). These groups were further recoded into 2 groups: groups with mixed users (n = 102) and groups with either all inexperienced users or all experienced users (n = 37). Multivariate repeated measures analyses found no significant 3- or 2-way interaction effect of group composition based on prior computer experience, learning method, and time of measurement (learning condition × prior computer experience composition: P = .40; time × prior computer experience composition: P = .20; time × learning condition × prior computer experience composition: P = .53). Further, no significant main effect of group composition based on prior computer experience was found (P = .54). Hypothesis 3 was not supported.

Learning Preferences

To examine the potential impact of attitude toward the individualistic versus collaborative learning method (learning preference), participants were recoded into a "matched" group, in which participants' learning preferences matched their collaborative versus individualistic group assignments, and a "no-match" group, where the preference and the group assignment did not match. Multivariate repeated measures analyses found neither an interaction effect among learning preference matching, learning method, and time of measurement (P = .18) nor a significant main effect of the learning preference matching factor (P = .41).

Discussion

Older adults are in great need of health literacy interventions, given that their needs for health information and services are typically high [17,60-62] and yet their health literacy levels are low [8]. Due to age-related changes in social environments and individual abilities [63], interventions that target younger age groups are unlikely to reach or have similar impact on older adults. Further complicating the situation is that the requirement for health literacy skills is a moving target, particularly in the context of eHealth becoming increasingly prominent in contemporary society [9,10]. As Norman [23] correctly points out, as technology changes, so do the requirements for health literacy skills.

Impact of the Overall Intervention

This study aimed to generate new scientific knowledge about effective eHealth literacy interventions for the older population. The primary research question was "What impact might the intervention have on the learning outcomes?" The analyses revealed that computer knowledge, Web knowledge, procedural skills, and eHealth literacy efficacy all improved significantly from pre to post intervention (P < .001 in all cases). The effect sizes of these improvements ranged from 0.88 to 2.25, suggesting that the magnitude of these improvements was large [64]. What these effect sizes mean is that, for instance, with respect to improvements in eHealth literacy efficacy, an effect size of 2.25 meant a learner increased from the 50th percentile on the pre test to the 99th percentile on the post test on this measure. These results strongly support the magnitude of the effects of the intervention. Further adding to the strength of these positive results is that the statistical power of these measures was strong: it reached 1.00 even at the alpha = .01 level in all cases.

These findings are even more impressive when interpreted in the context of the literature showing that "effect sizes of 0.8 are rare for any [learning] intervention and require truly impressive gains" ([31] p 224). Also, as summarized in several meta-analyses, the effect sizes of prior collaborative learning interventions (that focused on younger learners in formal educational settings) ranged from 0.29 to 0.70 [33,65,66]. In the present study, the effect size of the intervention on all knowledge, skill, and efficacy measures was greater than 0.8, suggesting the intervention has indeed resulted in "truly impressive gains." These results provide strong support that the intervention, regardless of the specific learning method used, was effective in improving older adults' eHealth literacy.

Across the 2 experimental conditions, participants had overwhelmingly positive views of the intervention. Notable percentages of participants also reported changes in various aspects of participation in their own health care. These findings suggest the health information these participants obtained from the intervention had affected their health behavior and decision making, which is a key component of eHealth literacy [14]. These findings further suggest that the intervention, regardless of the specific learning method used, was effective in improving older adults' eHealth literacy. These findings are particularly meaningful in the context of the contemporary health care system increasingly promoting shared medical decision making, where patients are expected to participate more in their own health care [60,67-71].

An important reason for the effectiveness of the intervention tested in this study was that it fully incorporated the key elements of successful computer training for older adults [24-30], as outlined in the Introduction section above. These key elements were proved effective in prior *e*HiLL studies [4-6,24]. A unique aspect of this study was to compare the effectiveness of the collaborative and individualistic learning methods built into the intervention. The analyses yielded interesting findings, as discussed in the next subsection.

Collaborative Versus Individualistic Learning Method

The analyses found neither an interaction effect of the learning method and time of measurement nor a main effect of the learning method on any of the outcome measures, suggesting the collaborative and individualistic learning methods did not differ in their relative effects. This finding deserves careful

XSL•FO
RenderX

consideration. As reviewed above, the superiority of collaborative learning over individualistic learning, as is well documented in the literature, is based on studies of younger learners in formal educational settings [33,34]. It is possible that collaborative learning may simply not work as well for the older population in informal settings. If this is the case, then this study contributes to the literature by identifying some key limits of the social interdependency theory.

In particular, the sample of this study consisted of primarily inexperienced computer users: less than 9% of participants used computers every day, while 35% of participants had never used a computer before (participants who used computers less often than every 2–3 days made up 72% of the study sample). Engaging in collaborative learning might have been too challenging for most participants who had to focus on their own activities with little attention to spare to interact with and help others, which is in line with the findings of prior research [25,28]. Thus, the previously widely reported advantages of collaborative learning over individualistic learning may not be easily realized among individuals who have limited prior computer experience.

Another way to look at this matter, however, is that perhaps different collaborative learning strategies could be used to better promote collaboration among inexperienced computer users. The collaborative learning strategies used in this study were carefully developed based on prior research on younger adult learning in formal educational settings [33,34] and modified to accommodate the older population in informal settings (see Table 4). In reflection, however, there might not have been sufficient consideration for the largely inexperienced user composition of the study sample.

For logistical reasons, multiple instructors were hired on an hourly basis to provide the instructions. It is likely that individual differences in these instructors (eg, teaching style, commitment, personality) may have affected the learning outcomes. One indication of this possibility is that the measure of the instructor's teaching *approached* a statistically significant difference between the 2 experimental conditions ($P = .07$). Participants in the individualistic learning condition expressed more positive views of the instructor's teaching than those in the collaborative learning condition. It is possible that this factor might have helped at least partially offset any hypothesized advantage of collaborative learning over individualistic learning.

The ceiling effect might have also affected the results: due to the positive impact of either version of the intervention, it may have been difficult to differentiate between the relative effects of the collaborative versus individualistic versions of the intervention. A possible solution for future research is to make the knowledge and skill tests more challenging so that the measures can be more sensitive to potential differences in learning outcomes.

Group Composition

As reviewed above, the literature suggests that group composition factors (based on gender, prior familiarity with peers, and prior computer experience) may affect the learning outcome. This study, however, found neither an interaction effect of any of these group composition factors with the collaborative versus individualistic learning method nor a main effect of any of these group composition factors. These findings are in line with those of 2 earlier *e*HiLL studies [5,6]. Replicating the same findings in these 3 independent studies, which differed in multiple ways (eg, intervention duration, instructional materials, procedures, and participants), lends some support to the generalizability of these findings. Note, though, that the study samples of these 3 independent *e*HiLL studies were similar: in each study the majority of the participants were women, unfamiliar with their study peers, and inexperienced computer users. While these findings might be generalized to populations with similar characteristics, they may not be so to populations with different characteristics. Also, as discussed above, if there were insufficient strategies to fully promote collaborative learning, then the potential effect of group composition might have also been affected.

Participants who completed both the pre and post testing reported *less* prior use of computers than those who completed only the pre testing ($P < .01$), suggesting that participants who used computers more often were more likely to drop out. One possible reason is that the intervention, by design, started from basic computer terms and increased in complexity gradually (see Table 3). Yet this might not have been made clear to the participants and some of them, after the first session, might have gotten the impression that the class was too "basic" for them and thus left. In future research, it will be necessary to fully communicate to participants, during the very first session, all the topics that will be covered in the remaining sessions.

Practical Implications

As in the earlier *e*HiLL interventions [4-6,24], the intervention tested in the present study also involved productive partnerships among local public libraries, a library and information science academic program at a state university, and the NIH. These local, state, and federal organizations bring complementary resources to the project and, in doing so, help each organization to achieve its mission [4]. The local public libraries provide the facility and staff support for the project, helping the libraries better serve socially, economically, and technologically underserved library patrons. The library and information science academic program provides the human resources through well-trained and dedicated faculty and graduate students and, in doing so, better achieves its research and educational missions. The NIH provides reliable online health information resources, and its involvement in this study helps promote the use of these resources. Tapping into these well-established public infrastructures ensures the intervention's capacity for scaling up (eg, it can be easily rolled out to other communities across the country to improve older adults' eHealth literacy).

Limitations and Future Directions

This study has some limitations. First, the sample was a convenience one, consisting of mostly African American/black people (64%), women (69%), and inexperienced computer users (72%). The findings may not be representative of the older population as a whole and should not be generalized without caution. Future research will benefit from examining the issues in a representative sample.

XSL•FO
RenderX

Second, potential confounding factors might have affected the relative effects of collaborative versus individualistic learning. These include insufficient consideration for the inexperienced user composition of the sample, instructor differences, and the ceiling effect. It will be necessary in future research to test strategies that can more fully promote collaboration among participants who have limited computer experience, control for instructor differences, and use more sensitive measurements to eliminate a potential ceiling effect.

Third, eHealth literacy skills may involve different levels of skills, with some skills being relatively easier to obtain and others requiring more effort. In future research, it will be necessary to develop more refined measures to assess changes in skills on different levels (eg, skills in not only finding a particular health topic on the NIHSeniorHealth.gov site but also determining the quality of information on any health website).

Fourth, this experimental study did not have a qualitative component. Future research can include qualitative data collection and analysis, which may generate additional insights into the learning process and the relative effects of collaborative versus individualistic learning.

Fifth, this study did not include measures of participants' potential practice of the skills outside of the intervention (eg, at home or other locations). Future research may include these measures to determine whether and how outside practice might affect the effects of the intervention.

Finally, changes in participation in own health care were measured in this study by self-report and at only one time point (post intervention) with no follow-up beyond the intervention period. In future research, it will be necessary to add more objective measures (eg, physicians' reports) and measure this variable over time at multiple time points.

Conclusions

The findings of this experimental study contribute to the literatures on adult learning, social interdependence theory, and health literacy. This study used both objective and self-report measures. The findings from both types of measures are consistent and, together, provide strong evidence that the eHealth literacy intervention tested in this study, regardless of the specific learning method used, significantly improved knowledge, skills, and efficacy. Participants also had overwhelmingly positive attitudes toward the intervention. Participants reported changes in participation in their own health care as a result of the intervention, further supporting the effectiveness of the intervention on improving older adults' eHealth literacy. Collaborative learning did not differ from individualistic learning in affecting the learning outcomes, suggesting that the previously widely reported advantages of collaborative learning over individualistic learning may not be easily applied to the older population in informal settings, though several confounding factors might have contributed to this finding. Further research is necessary before a more firm conclusion can be drawn. Finally, group composition based on gender, familiarity with peers, and prior computer experience demonstrated no significant interaction or main effect on the learning outcomes.

The study addressed an important social problem: the health "illiteracy" problem among older adults, particularly those who have low incomes, limited education, limited prior computer/Internet experience, and/or belong to ethnic minority groups. The findings of this study contribute to scientific knowledge by advancing theory in older adult learning, particularly the generalizability and application of the collaborative versus individualistic learning method to the older population in an informal setting, and the use of these learning methods as an effective eHealth literacy intervention. By focusing the content of learning on eHealth literacy knowledge and skills, this study broadens current understanding of the health literacy concept and interventions to address the increasing importance of technology in health care. By developing and testing the effectiveness of an eHealth literacy intervention, this study shapes this newly emerging component of health literacy (ie, eHealth literacy) that has increasing significance in contemporary health care.

This study broadens current paradigms in health literacy by using concepts and approaches novel to the field of health literacy. First, while health literacy has been promoted as a lifelong learning process [22], little attention has been paid to examining the relative benefits of different instructional methods (eg, individualistic versus collaborative learning) on older adults' learning of health literacy knowledge and skills. Second, while some prior interventions have involved the use of computers, their primary approach is presenting medical materials on a specific topic through a specially designed localized computer-based system [21]. While such an approach has its advantages (eg, targeting a very specific problem), it also has limitations: it requires extensive resources (to develop and update), and the knowledge and skills learned through this approach are often difficult to generalize to other areas or computer systems. In contrast, in the present study, we used the high-quality Internet health information resources maintained and updated by the NIH that focus on a broad range of medical knowledge, and provide training to improve general knowledge about and skills in finding information on any health topic that might be of interest to an individual—and from a common computer system. Thus, this *e*HiLL intervention is cost effective and easily transferrable.

Acknowledgments

The author thanks the administrators and staff of the Prince George's County Memorial Library System for facilitating the on-site recruitment and logistics, the instructors for carrying out the instructions, graduate assistant Ivan Watkins for scheduling and administering the data collection instruments, and graduate assistant Man Huang for administering the data collection instruments and data entry. Special thanks to the reviewers and the editor for their constructive and helpful feedback on earlier versions of

XSL•FO
RenderX

this manuscript. The Electronic Health Information for Lifelong Learners (eHiLL) research project is funded by a Faculty Early Career Development Award from the Institute of Museum and Library Services (RE-04-09-0067-09; PI: Bo Xie) and an R01 grant from the National Institute on Aging and the Office of Behavioral and Social Sciences Research (OBSSR) in the Office of the Director, National Institutes of Health (1R01AG041284-01; PI: Bo Xie).

Conflicts of Interest

None declared

References

1. US Department of Health and Human Services. Healthy People 2010: Understanding and Improving Health. 2nd edition. Washington, DC: US Government Printing Office; 2000.
2. Stvilia B, Mon L, Yong JY. A model for online consumer health information quality. J Am Soc Inf Sci Technol 2009;60(9):1781-1791. [doi: 10.1002/asi.21115]
3. Lustria MLA. Can interactivity make a difference? Effects of interactivity on the comprehension of and attitudes toward online health content. J Am Soc Inf Sci Technol 2007;58(6):766-776. [doi: 10.1002/asi.20557]
4. Xie B, Bugg JM. Public library computer training for older adults to access high-quality Internet health information. Libr Inf Sci Res 2009;31(3):155-162. [doi: 10.1016/j.lisr.2009.03.004] [Medline: 20161649]
5. Xie B. Older adults, e-health literacy, and collaborative learning: an experimental study. J Am Soc Inf Sci Technol 2011;62(5):933-946. [doi: 10.1002/asi.21507]
6. Xie B. Experimenting on the impact of learning methods and information presentation channels on older adults' e-health literacy. J Am Soc Inf Sci Technol 2011;62(9):1797-1807. [doi: 10.1002/asi.21575]
7. Nielsen-Bohlman L, Panzer AM, Kindig DA, Institute of Medicine. Health Literacy: A Prescription to End Confusion. Washington, DC: National Academies Press; 2004.
8. Kutner M, Greenberg E, Jin Y, Paulsen C. National Center for Education Statistics. 2006. The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy (NCES 2006-483) URL: http://nces.ed.gov/pubs2006/2006483.pdf [accessed 2011-08-17] [WebCite Cache ID 610sP11KU]
9. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth?: a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [Medline: 15829471] [doi: 10.2196/jmir.7.1.e1]
10. Hernandez LM. Health Literacy, eHealth, and Communication: Putting the Consumer First: Workshop Summary. Washington, DC: National Academies Press; 2009.
11. Bylund CL, Sabee CM, Imes RS, Sanford AA. Exploration of the construct of reliance among patients who talk with their providers about internet information. J Health Commun 2007;12(1):17-28. [doi: 10.1080/10810730601091318] [Medline: 17365346]
12. Fox S. E-patients with a disability or chronic disease. Washington, DC: Pew Internet & American Life Project; 2007 Oct 08. URL: http://www.pewinternet.org/~/media//Files/Reports/2007/EPatients_Chronic_Conditions_2007.pdf.pdf [accessed 2011-10-29] [WebCite Cache ID 62o2c4anF]
13. Huntington P, Nicholas D, Jamali HR, Russell C. Health information for the consumer: NHS vs the BBC. Aslib Proc 2007;59(1):46-67. [doi: 10.1108/00012530710725205]
14. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
15. Jones S, Fox S. Generations online in 2009. Washington, DC: Pew Internet & American Life Project; 2009 Jan 28. URL: http://www.pewinternet.org/~/media//Files/Reports/2009/PIP_Generations_2009.pdf [accessed 2011-10-29] [WebCite Cache ID 62o2oy9hv]
16. Rideout V, Neuman T, Kitchman M, Brodie M. Kaiser Family Foundation. 2005 Jan. e-Health and the Elderly: How Seniors Use the Internet for Health Information URL: http://www.kff.org/entmedia/upload/e-Health-and-the-Elderly-How-Seniors-Use-the-Internet-for-Health-Information-Key-Findings-From-a-National-Survey-of-Older-Americans-Survey-Report.pdf [accessed 2011-10-29] [WebCite Cache ID 62o3DMlkI]
17. Xie B. Older adults, health information, and the Internet. Interactions 2008;15(4):44-46.
18. Mika VS, Kelly PJ, Price MA, Franquiz M, Villarreal R. The ABCs of health literacy. Fam Community Health 2005;28(4):351-357. [Medline: 16166862]
19. Parker R, Kreps GL. Library outreach: overcoming health literacy challenges. J Med Libr Assoc 2005 Oct;93(4 Suppl):S81-S85. [Medline: 16239962]
20. Andrus MR, Roth MT. Health literacy: a review. Pharmacotherapy 2002 Mar;22(3):282-302. [Medline: 11898888]
21. Schaefer CT. Integrated review of health literacy interventions. Orthop Nurs 2008;27(5):302-317. [doi: 10.1097/01.NOR.0000337283.55670.75] [Medline: 18832992]
22. Kickbusch I. Kickbusch Health Consult. 2004. Background Paper: Improving Health Literacy in the European Union: Towards a Europe of Informed and Active Health Citizens URL: http://www.ilonakickbusch.com/health-literacy/health-literacy-gastein.pdf [accessed 2011-08-17] [WebCite Cache ID 610sfpbYK]

XSL•FO
RenderX

23. Norman CD. Skills essential for eHealth. In: Hernandez LM, editor. Health Literacy, eHealth, and Communication: Putting the Consumer First: Workshop summary. Washington, DC: National Academies Press; 2009:10-14.
24. Xie B. Improving older adults' e-health literacy through computer training using NIH online resources. Libr Inf Sci Res 2012.
25. Mayhorn CB, Stronge AJ, McLaughlin AC, Rogers WA. Older adults, computer training, and the systems approach: a formula for success. Educ Gerontol 2004;30(3):185-203. [doi: 10.1080/03601270490272124]
26. Van Fleet C, Antell K. Creating cyber seniors: older adult learning and its implications for computer training. Public Libr 2002;41:149-155.
27. Jay GM, Willis SL. Influence of direct computer experience on older adults' attitudes toward computers. J Gerontol B Psychol Sci Soc Sci 1992;47(4):250-257.
28. Cody MJ, Dunn D, Hoppin S, Wendt P. Silver surfers: Training and evaluating Internet use among older adult learners. Commun Educ 1999;48(4):269-286.
29. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352. [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
30. Bean C. Meeting the challenge: training an aging population to use computers. Southeast Libr 2003;51(3):16-25.
31. Prince M. Does active learning work? A review of the research. J Eng Educ 2004;93(3):223-231.
32. Bonwell CC, Eison JA. Active Learning: Creating Excitement in the Classroom. ASHEERIC Higher Education Report Series No. 1. Washington, DC: Jossey-Bass; 1991.
33. Johnson DW, Johnson RT, Smith K. Cooperative learning returns to college: what evidence is there that it works? Change 1998;30:27-35.
34. Johnson DW, Johnson RT, Smith K. The state of cooperative learning in postsecondary and professional settings. Educ Psychol Rev 2007;19(1):15-29. [doi: 10.1007/s10648-006-9038-8]
35. Strough J, Margrett JA. Overview of the special section on collaborative cognition in later adulthood. Int J Behav Dev 2002;26(1):2-5. [doi: 10.1080/01650250143000300]
36. Elias JW, Wagster MV. Developing context and background underlying cognitive intervention/training studies in older populations. J Gerontol B Psychol Sci Soc Sci 2007 Jun;62 Spec No 1:5-10. [Medline: 17565160]
37. Meegan S, Berg CA. Contexts, functions, forms, and processes of collaborative everyday problem solving in older adulthood. Int J Behav Dev 2002;26(1):6-15. [doi: 10.1080/01650250143000283]
38. Fried LP, Carlson MC, Freedman M, Frick KD, Glass TA, Hill J, et al. A social model for health promotion for an aging population: initial evidence on the Experience Corps model. J Urban Health 2004;81(1):64-78. [doi: 10.1093/jurban/jth094] [Medline: 15047786]
39. Stine-Morrow EA, Parisi JM, Morrow DG, Park DC. The effects of an engaged lifestyle on cognitive vitality: a field experiment. Psychol Aging 2008 Dec;23(4):778-786. [doi: 10.1037/a0014341] [Medline: 19140649]
40. Andersson J, Ronnberg J. Recall suffers from collaboration: joint recall effects of friendship and task complexity. Appl Cogn Psychol 1995;9(3):199-211. [doi: 10.1002/acp.2350090303]
41. Andersson J, Ronnberg J. Collaboration and memory: effects of dyadic retrieval on different memory tasks. Appl Cogn Psychol 1996;10(2):171-181. [doi: 10.1002/(SICI)1099-0720(199604)]
42. Margrett JA, Willis SL. In-home cognitive training with older married couples: individual versus collaborative learning. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2006 Jun;13(2):173-195. [doi: 10.1080/138255890969285] [Medline: 16807197]
43. Gould O, Kurzman D, Dixon RA. Communication during prose recall conversations by young and old dyads. Discourse Process 1994;17(1):149-165. [doi: 10.1080/01638539409544863]
44. Maskit D, Hertz-Lazarowitz R. Adults in cooperative learning: effects of group size and group gender composition on group learning. 1986 Presented at: Annual Meeting of the American Educational Research Association; April 16-20, 1986; San Francisco, CA, USA.
45. Busch T. Gender, group composition, cooperation, and self-efficacy in computer studies. J Educ Comput Res 1996;15(2):125-135.
46. Underwood G, McCaffrey M, Underwood J. Gender differences in a cooperative computer-based language task. Educ Res 1990;13:21-39.
47. Margrett JA, Marsiske M. Gender differences in older adults' everyday cognitive collaboration. Int J Behav Dev 2002 Jan;26(1):45-59. [doi: 10.1080/01650250143000319] [Medline: 20657668]
48. Charness N, Kelley CL, Bosman EA, Mottram M. Word-processing training and retraining: effects of adult age, experience, and interface. Psychol Aging 2001 Mar;16(1):110-127. [Medline: 11302360]
49. Nair SN, Czaja SJ, Sharit J. A multilevel modeling approach to examining individual differences in skill acquisition for a computer-based task. J Gerontol B Psychol Sci Soc Sci 2007 Jun;62 Spec No 1:85-96. [Medline: 17565169]
50. Czaja SJ, Lee CC. The Internet and older adults: design challenges and opportunities. In: Charness N, Parks DC, Sabel BA, editors. Communication, Technology and Aging: Opportunities and Challenges for the Future. New York, NY: Springer; 2001:60-78.

XSL•FO
RenderX

51. Zandri E, Charness N. Training older and younger adults to use software. Educ Gerontol 1989;15(6):615-631. [doi: 10.1080/0380127890150606]
52. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
53. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.
54. Pace RC, Kuh GD. Indiana University Center for Postsecondary Research and Planning, School of Education. 1988. College Student Experiences Questionnaires URL: http://dpb.cornell.edu/documents/1000093.pdf [accessed 2011-10-29] [WebCite Cache ID 62o4A8Umw]
55. Fox S. Online health search 2006. Washington, DC: Pew Internet & American Life Project; 2006 Oct 29. URL: http://www.pewinternet.org/~/media//Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf [accessed 2011-08-17] [WebCite Cache ID 610snOh4j]
56. NIHSeniorHealth. National Institute on Aging. 2010 Oct 12. Helping Older Adults Search for Health Information Online: A Toolkit for Trainers URL: http://nihseniorhealth.gov/toolkit/toolkit.html [accessed 2011-08-19] [WebCite Cache ID 613iGk3mp]
57. Childs S. Judging the quality of Internet-based health information. Perform Meas Metr 2005;6(2):80-96. [doi: 10.1108/14678040510607803]
58. Kunst H, Groot D, Latthe PM, Latthe M, Khan KS. Accuracy of information on apparently credible websites: survey of five common health topics. BMJ 2002 Mar 9;324(7337):581-582 [FREE Full text] [Medline: 11884323]
59. Morrell RW, Dailey SR, Rousseau GK. Applying research: the NIHSeniorHealth.gov project. In: Charness N, Warner Schaie K, editors. Impact of Technology on Successful Aging. New York, NY: Springer; 2003:134-161.
60. Xie B. Older adults' health information wants in the internet age: implications for patient-provider relationships. J Health Commun 2009 Sep;14(6):510-524. [doi: 10.1080/10810730903089614] [Medline: 19731124]
61. Williamson K. Discovered by chance: the role of incidental information acquisition in an ecological model of information use. Libr Inf Sci Res 1998;20(1):23-40. [doi: 10.1016/j.lisr.2009.03.004]
62. Asla T, Williamson K, Mills J. The role of information in successful aging: the case for a research focus on the oldest old. Libr Inf Sci Res 2006;28(1):49-63. [doi: 10.1016/j.lisr.2005.11.005]
63. Birren J, Warner Schaie K. Handbook of the Psychology of Aging. 3rd edition. San Diego, CA: Academic Press; 1990.
64. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale, NJ: L Erlbaum Associates; 1988.
65. Johnson DW, Johnson RT, Smith KA. Active Learning: Cooperation in the College Classroom. Edina, MN: Interaction Book Co; 1991.
66. Springer L, Stanne M, Donovan S. Effects of small-group learning on undergraduates in science, mathematics, engineering and technology: a meta-analysis. Rev Educ Res 1999;69(1):21-52. [doi: 10.3102/00346543069001021]
67. Brody DS. The patient's role in clinical decision-making. Ann Intern Med 1980 Nov;93(5):718-722. [Medline: 7212484]
68. Jones JA, Phillips GM. Communicating With Your Doctor: Rx for Good Medical Care. Carbondale, IL: Southern Illinois University Press; 1988.
69. Ballard-Reisch DS. A model of participative decision making for physician-patient interaction. Health Commun 1990;2(2):91-104. [doi: 10.1207/s15327027hc0202_3]
70. McNutt RA. Shared medical decision making: problems, process, progress. JAMA 2004 Nov 24;292(20):2516-2518. [doi: 10.1001/jama.292.20.2516] [Medline: 15562133]
71. Benbassat J, Pilpel D, Tidhar M. Patients' preferences for participation in clinical decision making: a review of published surveys. Behav Med 1998;24(2):81-88. [doi: 10.1080/08964289809596384] [Medline: 9695899]

Abbreviations

ANOVA: analysis of variance
eHiLL: Electronic Health Information for Lifelong Learners

Edited by G Eysenbach; submitted 29.06.11; peer-reviewed by ML Lustria, D Kaufman; comments to author 04.08.11; revised version received 14.09.11; accepted 27.09.11; published 03.11.11.

Please cite as:
Xie B
Effects of an eHealth Literacy Intervention for Older Adults
J Med Internet Res 2011;13(4):e90
URL: http://www.jmir.org/2011/4/e90/
doi:10.2196/jmir.1880
PMID:22052161

XSL•FO
RenderX

©Bo Xie. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 03.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Does the eHealth Literacy Scale (eHEALS) Measure What it Intends to Measure? Validation of a Dutch Version of the eHEALS in Two Adult Populations

Rosalie van der Vaart[1], MS; Alexander JAM van Deursen[2], PhD; Constance HC Drossaert[1], PhD; Erik Taal[1], PhD; Jan AMG van Dijk[2], PhD; Mart AFJ van de Laar[1,3], MD, PhD

[1]Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

[2]Department of Media, Communication and Organization, University of Twente, Enschede, Netherlands

[3]Arthritis Centre Twente, Enschede, Netherlands

Corresponding Author:
Rosalie van der Vaart, MS
Department of Psychology, Health and Technology
University of Twente
Citadel H423
Drienerlolaan 5, PO Box 217
Enschede, 7500AE
Netherlands
Phone: 31 53 489 6099
Fax: 31 53 489 2388
Email: r.vandervaart@utwente.nl

Abstract

Background: The Internet increases the availability of health information, which consequently expands the amount of skills that health care consumers must have to obtain and evaluate health information. Norman and Skinner in 2006 developed an 8-item self-report eHealth literacy scale to measure these skills: the eHealth Literacy Scale (eHEALS). This instrument has been available only in English and there are no data on its validity.

Objectives: The objective of our study was to assess the internal consistency and the construct and predictive validity of a Dutch translation of the eHEALS in two populations.

Methods: We examined the translated scale in a sample of patients with rheumatic diseases (n = 189; study 1) and in a stratified sample of the Dutch population (n = 88; study 2). We determined Cronbach alpha coefficients and analyzed the principal components. Convergent validity was determined by studying correlations with age, education, and current (health-related) Internet use. Furthermore, in study 2 we assessed the predictive validity of the instrument by comparing scores on the eHEALS with an actual performance test.

Results: The internal consistency of the scale was sufficient: alpha = .93 in study 1 and alpha = .92 in study 2. In both studies the 8 items loaded on 1 single component (respectively 67% and 63% of variance). Correlations between eHEALS and age and education were not found. Significant, though weak, correlations were found between the eHEALS and quantity of Internet use (r = .24, P = .001 and r = .24, P = .02, respectively). Contrary to expectations, correlations between the eHEALS and successfully completed tasks on a performance test were weak and nonsignificant: r = .18 (P = .09). The t tests showed no significant differences in scores on the eHEALS between participants who scored below and above median scores of the performance test.

Conclusions: The eHEALS was assessed as unidimensional in a principal component analysis and the internal consistency of the scale was high, which makes the reliability adequate. However, findings suggest that the validity of the eHEALS instrument requires further study, since the relationship with Internet use was weak and expected relationships with age, education, and actual performance were not significant. Further research to develop a self-report instrument with high correlations with people's actual eHealth literacy skills is warranted.

(J Med Internet Res 2011;13(4):e86) doi:10.2196/jmir.1840

XSL•FO
RenderX

KEYWORDS

e-health, literacy, internet, online, skills, health care, information

Introduction

Although a large supply of health information is available to educate and empower people, many lack the capability to use this information for their own benefit [1]. This capability is set out in the concept health literacy, which refers to the ability to read, understand, and communicate about health information to make proper health decisions [2]. In the Netherlands, 11% of the population has low literacy levels, according to the International Adult Literacy Survey [3], and it is assumed that the number of people who have limited levels of health literacy is even higher [4]. In other developed countries this problem is present to the same or worse extent [5,6]. These low levels of (health) literacy are worrisome, since health care is changing, and patients are increasingly expected to be involved in treatment, in health decisions, and in self-management of their disease [7]. As a result, there is an increasing gap between the needed level of health literacy to participate in their own health care, and the actual health literacy level of many patients. Consequently, low levels of health literacy might negatively influence health outcomes, success of treatment, and medical costs [8-10].

Online Health Information

With the increased diffusion of the Internet among households, the accessibility to relevant health information for the public has increased spectacularly. Controversially, this might also further enlarge the existing differences in health knowledge and access to care [11,12]. After all, collecting information through the Internet is different from collecting information through books and leaflets, and it requires specific skills [13-15]. For example, consumers should be able to use the computer, to navigate their way through the Internet, and to judge the large amount of information in terms of personal relevance, credibility, and accuracy [16]. Because the Internet and its impact keep growing, computer and Internet literacy are becoming an important addition to traditional health literacy skills [17]. Therefore, to get a complete overview of people's skills to obtain and use health information, we should measure eHealth literacy [11,14,18].

Insight into people's literacy skills is required to properly deploy guidelines, strategies, and interventions to offer information on different levels and in different formats. This is essential to make health information available and understandable to everyone who needs it [19].

Measurement of (e)Health Literacy

To measure health literacy levels, the Rapid Estimate of Adult Literacy in Medicine (REALM) [20] and the Test of Functional Health Literacy in Adults (TOFLA) [21] are often used. Both these instruments measure functional health literacy, which implies reading skills and, to some extent, numeracy. Other instruments that tend to measure a broader spectrum of health literacy skills have recently been developed—for example, the Newest Vital Sign [22], the functional, communicative, and critical health literacy scales assessment by Ishikawa et al [23], and the Health Literacy Skills Instrument by McCormack [24]. For the measurement of health-related Internet skills, fewer instruments are available. Recently, Van Deursen and Van Dijk [25,26] proposed an in-depth definition of Internet skills, consisting of operational skills (basic skills to use the Internet), formal skills (navigation and orientation), information skills (finding information), and strategic skills (using the information for personal benefits). This definition derives from the essential combination in eHealth literacy of both technical aspects, related to the use of the Internet, and substantive aspects, related to the content provided by the Internet. The definition contains gradients of difficulty, while the four skills have a sequential and conditional nature [27]. The combination of these four Internet skills illustrates that the application of operational and formal skills alone is not sufficient when using the Internet. On the other hand, using information and strategic skills often depends on the presence of operational and formal skills to obtain information in the first place. All four types of skills can be measured in a series of performance tests in which participants are asked to complete assignments on the Internet (see Multimedia Appendix 1). While this is a valuable method to assess (health-related) Internet skills, it is also quite demanding, costly, and time consuming, which makes it a rather inefficient instrument to use for (clinical) practice and research purposes. Therefore, an easy-to-administer self-assessment instrument that combines the measurement of computer skills with health literacy skills is needed. To our knowledge, the only instrument available that claims to measure the health-related Internet skills of the general Internet user is the eHealth Literacy Scale (eHEALS) by Norman and Skinner [28].

The eHEALS

The eHEALS is an 8-item scale that tends to measure perceived skills at finding, evaluating, and applying electronic health information to health problems [28]. The instrument proved to be a reliable and easy-to-use self-report tool, and has been used in some studies [29,30]. The scale is based on a model that distinguishes between six types of literacy skills: traditional literacy, health literacy, information literacy, scientific literacy, computer literacy, and media literacy [31]. Accordingly, the eHEALS aims to measure a broad overview of literacy skills, which might make it a potential instrument to assess the effects of eHealth literacy-tailored strategies to deliver online information and applications. However, the eHEALS has until now been available only in English and, to our knowledge, there are no data on its validity. Therefore, the aim of the present study was to examine the reliability and the construct and predictive validity of a Dutch version of the eHEALS.

Methods

Two populations were studied, one containing patients with rheumatic diseases (study 1) and one containing a stratified sample of the general Dutch population (study 2). Because there are no other instruments that measure eHealth literacy, we

XSL•FO
RenderX

measured convergent validity using the associated items age, education, and (health-related) Internet use. Predictive validity was measured by comparison with actual performance on various health-related Internet tasks [32]. Study 1 was originally designed to gain insight into patients' needs and wishes regarding a Web-based rheumatology patient portal and comprised a survey to measure age, education, general Internet use, health-related Internet use, and the eHEALS [33]. Study 2 was originally meant to gain insight into peoples' Internet skills and comprised a survey to measure age, education, Internet use, and the eHEALS, plus a series of assignments on an Internet-connected personal computer [32].

Study 1

Population

A random sample of patients with rheumatic diseases was selected from the patient database of the rheumatology clinic of Medisch Spectrum Twente, Enschede, the Netherlands. A total of 496 patients were sent a personal invitation letter and a paper-and-pencil questionnaire by their treating rheumatologists. Patients expected to experience difficulty in completing the survey (e.g. because of significant cognitive impairment or illiteracy) were excluded a priori by their treating rheumatologists. The invitation letter explained the purpose of the study, the use of data, the voluntary nature, and the anonymity of the participant; therefore, returned questionnaires could be presumed to provide consent. A reminder was sent to those who did not respond within 2 weeks. According to local regulations in the Netherlands (Medical Research [Human Subjects] Act) the study did not need approval of the ethical review board; only (nonintervention) studies with a high burden for patients have to be reviewed. For this study, patients who indicated in the questionnaire that they did not have access to the Internet were excluded.

Instruments

The questionnaire assessed the following: (1) gender, age, and education level, (2) general and health-related Internet use, and (3) the eHEALS. General Internet use was measured by 2 items: 1 yes/no item measuring access to the Internet, and 1 item on quantity of Internet use with answer options on a 5-point Likert scale ranging from "(almost) never" to "(almost) every day." Health-related Internet use was measured with 8 items on quantity of use of different kinds of health-related information. Each item could be answered on a 4-point Likert scale ranging from "never" to "regularly" (see Table 1 for a complete overview of topics). The original items of the eHEALS were translated into Dutch with forward and backward translation, according to World Health Organization guidelines [34]. The eHEALS contains 8 items, measured with a 5-point Likert scale with response options ranging from "strongly disagree" to "strongly agree." Total scores of the eHEALS are summed to range from 8 to 40, with higher scores representing higher self-perceived eHealth literacy. The original version of the eHEALS can be found in Table 2. The whole survey instrument was pretested with 6 participants. Minor revisions were made in formulation and layout according to the received remarks and recommendations.

Study 2

Population

A sample of 88 participants was recruited by randomly dialing telephone numbers in cities and villages in the region of Twente. A stratified sampling method was used to gain equal categories in gender, age, and education. When respondents indicated they were willing to participate, their contact and email address were recorded and a time for the research session was scheduled. All research sessions were scheduled at the University of Twente, which was an unfamiliar environment to all participants. Respondents received a follow-up letter in the mail for confirmation, and the day before the study respondents were reminded of the session by telephone. Respondents were awarded €25 for their participation.

Instruments

The sessions lasted approximately 1.5 hours and started off with a short questionnaire that assessed (1) gender, year of birth, and education level, (2) general Internet use, and (3) the eHEALS. General Internet use was measured with 3 items: 1 yes/no item measuring access to the Internet, 1 item measuring amount of Internet use in hours per week, and 1 item on Internet experience in years.

Subsequently, participants had to complete a performance test, which contained nine health-related assignments, based on the four defined Internet skills. Two assignments (consisting of eight tasks) were used to measure operational Internet skills (e.g. open a health website, save a file, or add a website to the Favorites menu), two assignments (consisting four tasks) were used to measure formal Internet skills (e.g. navigate different health-related menu and website designs, and surf between different websites), three assignments were used to measure information Internet skills (find health-related information on the Internet), and two assignments were used to measure strategic Internet skills (e.g. extract information from different sources, and make decisions based on the information found). The assignments were generated by a team of researchers that made a conscious effort to include only tasks that were accessible and relevant to the general user population (e.g. find the Web address of a health clinic, or search for information on vitamins). All assignments were pilot tested with 12 participants to ensure comprehensibility and applicability. Assignments were administered in a sequence of increasing difficulty, as indicated in Multimedia Appendix 1. During the assignment completion, participants themselves decided when they were finished or wanted to give up on an assignment. Completion of the tasks, successful and unsuccessful, was directly noted during the sessions. Tasks were assumed successful if the right answer was given within an ample time period, determined in the pilot tests. To execute the assignments, participants used a keyboard, a mouse, and a 17-inch monitor. The personal computer was connected to the Internet on a high-speed university network and was programmed with the three most popular Internet browsers (Internet Explorer, Mozilla Firefox, and Google Chrome). This allowed the participants to replicate their regular Internet use. No default page was set on the browsers and all the assignments started with a blank page. To ensure that participants were not influenced by a previous user's actions,

XSL•FO
RenderX

the browser was reset after each session by removing temporary files, cookies, and favorites. In addition, downloaded files, history, forms, and passwords were removed and the laptop was rebooted.

Analysis

Data were analyzed using SPSS version 17.0 for Windows (IBM Corporation, Somers, NY, USA) in both studies. Cronbach alpha served as a measure of internal consistency, reflecting the (weighted) average correlation of items within the scale [35]. In general, a Cronbach alpha of .7 to .8 is regarded as satisfactory for scales to be used as research tools [36]. Principal component analysis was performed to examine the 1-factor structure of the scale. Factor loadings in excess of .71 were considered excellent, .63 very good, and .55 good [37].

Distributional properties of the eHEALS were further inspected to examine the normality of the total scores and to identify floor and ceiling effects. Skewness and kurtosis values between ±1 were assumed to indicate no or slight nonnormality. Floor or ceiling effects were considered to be present if >15% of the participants scored the worst or the best possible score on the eHEALS [38].

Evidence for convergent validity was determined by studying Spearman correlations between total mean scores on the eHEALS and age, education level, quantity of Internet use, and sum scores of health-related Internet use. Based on previous studies on regular health literacy, we hypothesized negative correlations with age and positive correlations with education and (health-related) Internet use [9,11,39]. A coefficient magnitude of at least .4 was taken as evidence of convergent validity [40]. The predictive validity of the instrument was assessed by comparing the total mean scores on the eHEALS with the scores on the actual performance test in study 2, using Spearman correlations. The scores on the eHEALS were first related to the total number of successfully completed tasks. Second, the scores on the eHEALS were related to the amount of completed tasks per skill (operational, formal, information and strategic). A coefficient magnitude of at least .4 was taken as evidence of predictive validity. We used *t* tests on each skill to investigate whether participants who performed below and above the median score of successfully completed assignments significantly differed on the eHEALS. Two-tailed *P* values less than .05 were considered significant.

Results

Study 1

Participants

Of the 496 invitations sent out, 12 were returned undeliverable. In total, 227 of 484 questionnaires were returned (47%); 189 of these 227 participants had Internet access and completed the eHEALS (83%). Participant characteristics and Internet use are shown in Table 1. Included respondents used the Internet daily or several days a week. Responders and nonresponders did not differ on gender, but nonresponders were on average 5 years younger, with a mean age of 47 years ($P < .001$).

XSL•FO
RenderX

Table 1. Participants' self-reported sociodemographics and (health-related) Internet use

	Study 1 (n = 189) n (%)	Study 2 (n = 88) n (%)
Gender		
Male	119 (63)	45 (51)
Female	70 (37)	43 (49)
Mean (SD) age (years)	52 (11)	43 (18)
Education level		
Low	38 (20)	25 (28)
Middle	102 (54)	32 (36)
High	46 (24)	31 (35)
Unknown	3 (2)	
Amount of Internet usage		
(almost) Every day	117 (62)	_[a]
Several days a week	34 (18)	
About 1 day a week	15 (8)	
Less than 1 day a week	9 (5)	
(almost) Never	12 (6)	
Unknown	2 (1)	
Mean (SD) Amount of Internet use (hours per week)	_[a]	12.2 (13.7)
Mean (SD) Internet experience (years)	_[a]	9.3 (4.3)
Number of respondents who have ever searched for information on:		
Diseases	159 (84)	_[a]
Healthy lifestyle	121 (64)	
Medication	95 (50)	
Treatments	122 (65)	
Care providers	69 (37)	
Patient organizations	67 (35)	
Law regulations related to health conditions	61 (34)	
Peer-support forums	45 (24)	

[a] Item was not measured in this study.

Distributional Properties

Total scores on the eHEALS were approximately normally distributed with a skewness of -.63. Floor and ceiling effects were acceptable, with no participants scoring the worst possible score (8), and 5 participants scoring the best possible score (40).

Reliability and Validity

The internal consistency of the eHEALS was alpha = .93. Unidimensionality of the scale was supported by principal component analysis (eigenvalue = 5.4, 67% of variance explained). The eigenvalue of the first component was 5 times larger than the eigenvalue of the second component (being 1.1). All items loaded high on this component, ranging from .74 to .85 (Table 2). The mean sum score of the scale was 28.2 (SD 5.9).

Table 3 shows the correlations between the scores on the eHEALS and the variables measured in both studies. Correlations with age (r = –.11, P= .13) and education (r = .09, P= .24) were not significant. A significant, though weak, positive correlation was found between the eHEALS and quantity of Internet use (r = .24, P = .001). Concerning health-related Internet use, the use of online information correlated weakly to moderately with the eHEALS with coefficients varying from .26 to .40 (P < .001).

XSL•FO
RenderX

Table 2. eHealth Literacy Scale (eHEALS) mean items scores, scale reliability, and principal component analysis

Item	Study 1		Study 2		Factor loading		item-total correlation[a]	
	Mean	SD	Mean	SD	Study 1	Study 2	Study 1	Study 2
1: I know what health resources are available on the Internet	3.6	0.83	3.4	0.86	.82	.77	.80	.70
2: I know where to find helpful health resources on the Internet	3.6	0.87	3.3	0.88	.85	.79	.84	.73
3: I know how to find helpful health resources on the Internet	3.7	0.81	3.5	0.94	.85	.86	.85	.72
4: I know how to use the Internet to answer my health questions	3.6	0.85	3.6	0.88	.83	.86	.83	.70
5: I know how to use the health information I find on the Internet to help me	3.5	0.88	3.4	.087	.84	.77	.85	.67
6: I have the skills I need to evaluate the health resources I find on the Internet	3.6	0.89	3.6	0.90	.82	.77	.84	.67
7: I can tell high-quality from low-quality health resources on the Internet	3.4	0.95	3.4	1.00	.80	.75	.82	.76
8: I feel confident in using information from the Internet to make health decisions	3.3	0.99	3.1	1.12	.74	.80	.78	.82
Mean (SD) sum score	28.2	5.9	27.6	5.9				
Eigenvalue first component	5.36		5.06					
Variance accounted for	67%		63%					
Cronbach alpha	.93		.92					

[a] All item-total correlations were significant at $P < .001$.

XSL•FO
RenderX

Table 3. Spearman correlations between scores on the eHealth Literacy Scale (eHEALS) and age, education, (health-related) Internet use, and Internet performance skills

	Study 1		Study 2	
	r	*P* value	*r*	*P* value
Sociodemographics				
Age	-.11	.13	-.08	.49
Education (1 = low, 2 = middle, 3 = high)	.09	.24	.13	.25
Amount of Internet usage	.24	.001	.24	.02
Health-related Internet use				
Information on diseases	.40	<.001	_[a]	
Healthy lifestyle	.28	<.001		
Medication	.29	<.001		
Treatments	.38	<.001		
Care providers	.30	<.001		
Patient organizations	.32	<.001		
Law regulations related to health conditions	.26	<.001		
Peer-support forums	.27	<.001		
Performance tasks				
Successfully completed tasks overall	_[a]		.18	.09
Operational			.12	.27
Formal			.19	.07
Information			.05	.62
Strategic			.11	.30

[a] Item was not measured in this study.

Study 2

Participants

Characteristics and Internet use of the 88 recruited participants in study 2 are shown in Table 1. Of all participants, 75 (85%) had home Internet access. The average years of Internet experience was 9.3 (SD 4.3) and average amount of Internet use was 12.2 hours a week (SD 13.7).

Performance Tests

Table 4 shows that the participants successfully completed an average of 73% (5.8/8) of the operational Internet skills tasks and an average of 73% (2.9/4) of the formal Internet skills tasks. Of the information Internet skills tasks, an average of 50% (1.5/3) was completed successfully and of the strategic Internet skills tasks, 35% (0.7/2). Only 28% (25/88) of the participants were able to successfully complete all operational skills tasks, 39% (34/88) completed all formal skills tasks, 13% (11/88) completed all information skills tasks, and 20% (18/88) completed both the strategic skill tasks. No participants exceeded the maximum amount of time they were given for the assignments. Participants who were not able to complete the assignment decided to give up on the assignment before the official end time had elapsed. More details on the results of the performance tests and the general consequences for health seekers and providers are discussed elsewhere [32].

Table 4. Overview of proportion of tasks successfully completed in performance tests

Internet skills (number of tasks)	Average task completion		
	Mean	SD	%
Operational tasks (8)	5.8	2.1	73
Formal tasks (4)	2.9	1.2	73
Information tasks (3)	1.5	0.9	50
Strategic tasks (2)	0.7	0.8	35

XSL•FO
RenderX

Distributional Properties

As in study 1, total scores on the eHEALS were approximately normally distributed with a slight skewness of –.80. Floor and ceiling effects were acceptable, with no participants scoring the worst possible score (8), and 4 participants scoring the best possible score (40).

Reliability and Validity

The internal consistency of the eHEALS was alpha = .92. All items loaded on 1 single component in this study as well (eigenvalue = 5.1, 63% of variance explained). The eigenvalue of the first component was 5.8 times larger than the eigenvalue of the second component (being .88). All items loaded high on this component, ranging from .75 to .86 (Table 2). The mean sum score of the scale was 27.6 (SD 5.9).

No significant correlations between the eHEALS and either age (r = -.08, P = .49) or education (r = .13, P = .25) were found (Table 3). A significant, though weak, correlation was found between the eHEALS and quantity of Internet use (r = .24, P = .02). The correlations between the eHEALS and actual performance for overall successfully completed tasks and the four skills separately were weak and nonsignificant (Table 3). Comparison on the four performance skills showed that the 50% of participants scoring above the median had a higher mean score on the eHEALS than the 50% of participants scoring below the median (Table 5). However, t tests showed that none of these differences were significant (Table 5).

Table 5. eHealth Literacy Scale (eHEALS) mean scores of participants scoring below and above median scores on performance tasks

Performance tasks	Mean	SD	*t* test	df	*P* value
Operational					
50% below median	3.38	0.85	–.998	80.33	.32
50% above median	3.53	0.59			
Formal					
50% below median	3.36	0.77	–1.47	77.38	.15
50% above median	3.59	0.67			
Information					
50% below median	3.43	0.69	–.26	81.37	.80
50% above median	3.47	0.80			
Strategic					
50% below median	3.38	0.74	–.79	81.55	.43
50% above median	3.51	0.74			

Discussion

The results of the two studies show that the eHEALS is unidimensional and has high internal consistency. Yet results of the validity tests showed that the eHEALS is not a valid measure of eHealth literacy.

With regard to the convergent validity, we hypothesized at least moderate positive correlations (r > .4) between scores on the eHEALS and education, and at least moderate negative correlations (r > .4) between the eHEALS and age. However, in both studies correlations between the eHEALS and either education or age were not significant. Although it should be noted that (selective) nonresponse might have had an influence, and that younger respondents (<30 years of age) were slightly underrepresented in study 1, we were surprised about the lack of these correlations, as various reviews have shown that these factors are the most predictive for (regular) health literacy [9,39]. In their study, Norman and Skinner [28] found no significant correlation between scores on eHEALS and age either, but in their study only adolescents in the age group of 13–21 years participated. To our knowledge, no other studies have examined the correlation between scores on eHEALS and age and education.

We hypothesized at least moderately positive correlations (r > .4) between scores on the eHEALS and quantity of Internet use, since it is reasoned that the amount of time spent on the Internet has a positive influence on eHealth literacy [11]. However, whereas the correlations between the scores on eHEALS and Internet experience were in the expected direction, they appeared to be weak in both of our studies. The correlations between eHEALS and health-related Internet use were weak but slightly higher, with Spearman correlation coefficients ranging from .26 to .40.

Concerning the predictive validity, the lack of significant correlations between the eHEALS and actual performance skills was surprising. Since the assignments used in study 2 were applicable to the general Internet user, one would at least expect some moderate correlations between the eHEALS scale and the performance results. Apparently, perceived skills (as obtained with eHEALS) do not predict actual performance (as measured in study 2). Previous investigations on general computer skills have also shown that people tend to overestimate their computer skills, which results in a gap between self-reported skills and practice when actual skills are measured [41,42]. Furthermore, the comparison of all participants who scored below and above median scores on the performance test did not show any significant differences on the eHEALS either. From this we can

XSL•FO
RenderX

conclude that the eHEALS does not have the power to distinguish between people with low health-related Internet skills and people with high health-related Internet skills. These results show that the eHEALS is not a valid instrument for assessing perceived health-related Internet skills.

We suggest a revision of the eHEALS, in a way that all four different skills are measured: (1) operational and (2) formal skills that measure practical use of computers and the Internet, and (3) information finding and (4) strategic skills that measure search strategies and skills to judge the found information. Also, questions might need to be formulated differently in order to prevent misunderstanding or differing interpretations. To this aim, qualitative research might provide more insight into the basis for participants' answers—for example, having people fill out the eHEALS with techniques such as cognitive interviewing or thinking-aloud methods [43,44]. When measuring al four different skills, we might obtain a more valid indication of eHealth literacy skills. This could also distinguish between what type of skills (groups of) people possess, after which proper implementation of interventions can bring about equal access to online health information for all subgroups.

Limitations

A limitation of both our studies is the voluntary basis on which participants were recruited. This could have caused a bias, because participants might already have been more interested in using the Internet and searching for information, which could have influenced the results. Concerning study 1, only patients with rheumatic diseases were invited to participate. Therefore, this study might not be representative for other chronic conditions, since patients with rheumatic diseases are on average somewhat older. Concerning study 2, because of the major labor intensity of performance tests and the very high travel costs of bringing participants nationwide to the university lab, it was not possible to test a random sample from the whole Dutch population. Although the study population size of 88 is not enough to generalize to the whole population, the applied quota sample for the categories of gender, age, and education hugely improved representativeness.

Conclusions

The eHEALS is found to be unidimensional, according to principal component analysis, and to be internally consistent, as assessed with Cronbach alpha, but its validity is questionable. Expected correlations between the eHEALS and peoples' use of the Internet were weak. Moreover, scores on the eHEALS did not correlate with age, education, and scores on performance tasks, and the eHEALS was not able to distinguish between people with high and low health-related Internet skills. Therefore, more research is needed in order to develop a self-report instrument that validly measures eHealth literacy skills. We suggest incorporation of operational, formal, information, and strategic Internet skills to measure all aspects of eHealth literacy.

Acknowledgments

The work of study 1 was supported by an unrestricted educational grant from Wyeth Pharmaceuticals, part of Pfizer Inc. This funding source had no involvement in data collection, analysis, or the preparation of this manuscript.

Conflicts of Interest

None declared

Multimedia Appendix 1

Performance test assignments.

[PDF File (Adobe PDF File), 53KB - jmir_v13i4e86_app1.pdf]

References

1. Mancuso JM. Assessment and measurement of health literacy: an integrative review of the literature. Nurs Health Sci 2009 Mar;11(1):77-89. [doi: 10.1111/j.1442-2018.2008.00408.x] [Medline: 19298313]
2. Baker DW. The meaning and the measure of health literacy. J Gen Intern Med 2006 Aug;21(8):878-883. [doi: 10.1111/j.1525-1497.2006.00540.x] [Medline: 16881951]
3. Organisation for Economic Co-operation and Development and Statistics Canada. 2000. Literacy in the Information Age: Final Report of the International Adult Literacy Survey URL: http://www.oecd.org/dataoecd/48/4/41529765.pdf [accessed 2011-03-23] [WebCite Cache ID 5xOuCqCN0]
4. Twickler TB, Hoogstraaten E, Reuwer AQ, Singels L, Stronks K, Essink-Bot ML. [Low literacy and limited health literacy require health care measures]. Ned Tijdschr Geneeskd 2009;153:A250. [Medline: 19785884]
5. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
6. Protheroe J, Nutbeam D, Rowlands G. Health literacy: a necessity for increasing participation in health care. Br J Gen Pract 2009 Oct;59(567):721-723. [doi: 10.3399/bjgp09X472584] [Medline: 19843420]
7. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. Health literacy: report of the Council on Scientific Affairs. JAMA 1999 Feb 10;281(6):552-557. [Medline: 10022112]

XSL•FO
RenderX

8. Jordan JE, Buchbinder R, Osborne RH. Conceptualising health literacy from the patient perspective. Patient Educ Couns 2010 Apr;79(1):36-42. [doi: 10.1016/j.pec.2009.10.001] [Medline: 19896320]
9. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005 Feb;20(2):175-184. [doi: 10.1111/j.1525-1497.2005.40245.x] [Medline: 15836552]
10. Nutbeam D, Kickbusch I. Advancing health literacy: a global challenge for the 21st century. Health Promot Int 2000;15(3):183-184. [doi: 10.1093/heapro/15.3.183]
11. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
12. Beacom AM, Newman SJ. Communicating health information to disadvantaged populations. Fam Community Health 2010;33(2):152-162. [doi: 10.1097/FCH.0b013e3181d59344] [Medline: 20216358]
13. Brandtweiner R, Donat E, Kerschbaum J. How to become a sophisticated user: a two-dimensional approach to e-literacy. New Media Soc 2010 Aug;12(5):813-833. [doi: 10.1177/1461444809349577]
14. Dominick GM, Friedman DB, Hoffman-Goetz L. Do we need to understand the technology to get to the science? A systematic review of the concept of computer literacy in preventive health programs. Health Educ J 2010 Dec;68(4):296-313. [doi: 10.1177/0017896909349289]
15. Van Deursen AJAM. Internet skills: Vital Assets in an Information Society. Enschede, Netherlands: University of Twente; 2010.
16. McCray AT. Promoting health literacy. J Am Med Inform Assoc 2005 Apr;12(2):152-163 [FREE Full text] [doi: 10.1197/jamia.M1687] [Medline: 15561782]
17. Coiro J. Reading comprehension on the Internet: Expanding our understanding of reading comprehension to encompass new literacy's. Reading Teacher 2003 Feb;56(5):458-464.
18. Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Saha S, Snyder A, et al. Variation in estimates of limited health literacy by assessment instruments and non-response bias. J Gen Intern Med 2010 Jul;25(7):675-681. [doi: 10.1007/s11606-010-1304-2] [Medline: 20224964]
19. Chang BL, Bakken S, Brown SS, Houston TK, Kreps GL, Kukafka R, et al. Bridging the digital divide: reaching vulnerable populations. J Am Med Inform Assoc 2004 Dec;11(6):448-457 [FREE Full text] [doi: 10.1197/jamia.M1535] [Medline: 15299002]
20. Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993 Jun;25(6):391-395. [Medline: 8349060]
21. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med 1995 Oct;10(10):537-541. [Medline: 8576769]
22. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med 2005;3(6):514-522 [FREE Full text] [doi: 10.1370/afm.405] [Medline: 16338915]
23. Ishikawa H, Takeuchi T, Yano E. Measuring functional, communicative, and critical health literacy among diabetic patients. Diabetes Care 2008 May;31(5):874-879 [FREE Full text] [doi: 10.2337/dc07-1932] [Medline: 18299446]
24. McCormack L, Bann C, Squiers L, Berkman ND, Squire C, Schillinger D, et al. Measuring health literacy: a pilot study of a new skills-based instrument. J Health Commun 2010;15 Suppl 2:51-71. [doi: 10.1080/10810730.2010.499987] [Medline: 20845193]
25. Van Deursen AJAM, Van Dijk JAGM. Measuring Internet skills. Int J Hum Comput Interact 2010 Oct;26(10):891-916. [doi: 10.1080/10447318.2010.496338]
26. Van Deursen AJAM, Van Dijk JAGM. Using the Internet: skill related problems in users' online behavior. Interact Comput 2009;21:393-402. [doi: 10.1016/j.intcom.2009.06.005]
27. Van Deursen AJAM, Van Dijk JAGM, Peters O. Rethinking Internet skills: the contribution Of gender, age, education, Internet experience, and hours online to medium- and content-related Internet skills. Poetics 2011;39:125-144. [doi: 10.1016/j.poetic.2011.02.001]
28. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
29. Brown CA, Dickson R. Healthcare students' e-literacy skills. J Allied Health 2010;39(3):179-184. [Medline: 21174023]
30. Robinson C, Graham J. Perceived Internet health literacy of HIV-positive people through the provision of a computer and Internet health education intervention. Health Info Libr J 2010 Dec;27(4):295-303. [doi: 10.1111/j.1471-1842.2010.00898.x] [Medline: 21050372]
31. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
32. van Deursen AJAM, van Dijk JAGM. Internet skills performance tests: are people ready for eHealth? J Med Internet Res 2011 Apr;13(2):e35 [FREE Full text] [doi: 10.2196/jmir.1581] [Medline: 21531690]
33. van der Vaart R, Drossaert CH, Taal E, van de Laar MAFJ. Patient preferences for a hospital-based rheumatology Interactive Health Communication Application and factors associated with these preferences. Rheumatology (Oxford) 2011 Sep;50(9):1618-1626. [doi: 10.1093/rheumatology/ker161] [Medline: 21551221]

XSL•FO
RenderX

34. World Health Organization. 2011. Process of Translation and Adaptation of Instruments URL: http://www.who.int/substance_abuse/research_tools/translation/en/index.html [accessed 2011-03-23] [WebCite Cache ID 5xOuj0r4a]
35. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16(3):297-334. [doi: 10.1007/BF02310555]
36. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford: Oxford University Press; 2003.
37. Comrey AL, Lee HB. A First Course in Factor Analysis. Hillsdale, NJ: Erlbaum; 1992.
38. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
39. Martin LT, Ruder T, Escarce JJ, Ghosh-Dastidar B, Sherman D, Elliott M, et al. Developing predictive models of health literacy. J Gen Intern Med 2009 Nov;24(11):1211-1216. [doi: 10.1007/s11606-009-1105-7] [Medline: 19760299]
40. Trochim WMK, Donnelly JP. The Research Methods Knowledge Base. Mason, OH: Atomic Dog; 2007.
41. Merritt K, Smith KD. Di Renzo Jr JC. An investigation of self-reported computer literacy: is it reliable? Issues Inf Syst 2005;6(1):289-295.
42. Van Vliet P, Kletke MG, Chakraborty G. The measurement of computer literacy: a comparison of self-appraisal and objective tests. Int J Hum Comput Stud 1994;40:835-857. [doi: 10.1006/ijhc.1994.1040]
43. Collins D. Pretesting survey instruments: an overview of cognitive methods. Qual Life Res 2003 May;12(3):229-238. [Medline: 12769135]
44. Willis GB. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA: Sage; 2005.

Edited by G Eysenbach; submitted 29.04.11; peer-reviewed by C Norman, L Donelle; comments to author 16.05.11; accepted 04.07.11; published 09.11.11.

Please cite as:
van der Vaart R, van Deursen AJAM, Drossaert CHC, Taal E, van Dijk JAMG, van de Laar MAFJ
Does the eHealth Literacy Scale (eHEALS) Measure What it Intends to Measure? Validation of a Dutch Version of the eHEALS in Two Adult Populations
J Med Internet Res 2011;13(4):e86
URL: http://www.jmir.org/2011/4/e86/
doi:10.2196/jmir.1840
PMID:22071338

©Rosalie van der Vaart, Alexander JAM van Deursen, Constance HC Drossaert, Erik Taal, Jan AMG van Dijk, Mart AFJ van de Laar. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 09.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Framework for Characterizing eHealth Literacy Demands and Barriers

Connie V Chan[1], PhD; David R Kaufman[1], PhD

Biomedical Informatics, Columbia University, New York, NY, United States

Corresponding Author:
David R Kaufman, PhD
Biomedical Informatics
Columbia University
Vanderbilt Clinic 5
622 West 168th St.
New York, NY, 10032
United States
Phone: 1 212 342 3461
Fax: 1 212 305 3302
Email: davek@dbmi.columbia.edu

Abstract

Background: Consumer eHealth interventions are of a growing importance in the individual management of health and health behaviors. However, a range of access, resources, and skills barriers prevent health care consumers from fully engaging in and benefiting from the spectrum of eHealth interventions. Consumers may engage in a range of eHealth tasks, such as participating in health discussion forums and entering information into a personal health record. eHealth literacy names a set of skills and knowledge that are essential for productive interactions with technology-based health tools, such as proficiency in information retrieval strategies, and communicating health concepts effectively.

Objective: We propose a theoretical and methodological framework for characterizing complexity of eHealth tasks, which can be used to diagnose and describe literacy barriers and inform the development of solution strategies.

Methods: We adapted and integrated two existing theoretical models relevant to the analysis of eHealth literacy into a single framework to systematically categorize and describe task demands and user performance on tasks needed by health care consumers in the information age. The method derived from the framework is applied to (1) code task demands using a cognitive task analysis, and (2) code user performance on tasks. The framework and method are applied to the analysis of a Web-based consumer eHealth task with information-seeking and decision-making demands. We present the results from the in-depth analysis of the task performance of a single user as well as of 20 users on the same task to illustrate both the detailed analysis and the aggregate measures obtained and potential analyses that can be performed using this method.

Results: The analysis shows that the framework can be used to classify task demands as well as the barriers encountered in user performance of the tasks. Our approach can be used to (1) characterize the challenges confronted by participants in performing the tasks, (2) determine the extent to which application of the framework to the cognitive task analysis can predict and explain the problems encountered by participants, and (3) inform revisions to the framework to increase accuracy of predictions.

Conclusions: The results of this illustrative application suggest that the framework is useful for characterizing task complexity and for diagnosing and explaining barriers encountered in task completion. The framework and analytic approach can be a potentially powerful generative research platform to inform development of rigorous eHealth examination and design instruments, such as to assess eHealth competence, to design and evaluate consumer eHealth tools, and to develop an eHealth curriculum.

(J Med Internet Res 2011;13(4):e94) doi:10.2196/jmir.1750

KEYWORDS

eHealth; health literacy; cognition; Bloom's taxonomy; cognitive task analysis; consumer health

XSL•FO
RenderX

Introduction

eHealth literacy names a set of skills and knowledge that are essential for productive interactions with technology-based health tools. The objective of this study was to explore how eHealth literacy can be systematically analyzed, measured, and quantified. We proposed a methodological and theoretical framework that systematically maps skill sets to successful performance of eHealth tasks. We employed a microanalytic strategy in which complex competencies can be broken down into constituent skills or local task demands. In our view, systematic understanding of the necessary competencies can inform development of targeted solution strategies to overcome skill- and knowledge-related barriers.

Background

The term eHealth refers to "health services and information delivered or enhanced through the Internet and related technologies" [1]. Consumer-oriented eHealth tools engage consumers in managing their own health care, communicating with providers and support networks, meeting their information needs, making health decisions, using patient education resources, and promoting healthy lifestyles [2-4]. Unfortunately, most of these eHealth tools have not been designed with the consideration of the needs and characteristics of diverse user groups. These tools may even increase the complexity of health care engagement for those lacking the prerequisite abilities [5].

Many different factors can inhibit consumers' meaningful use of eHealth tools, including environmental barriers [6], physical access barriers [7], resource-related barriers [8-11], and individual-level barriers [2,7,12,13]. Underserved and vulnerable populations face additional challenges that exacerbate these obstacles [14]. Different types of tools offer varied resources and functionalities, enabling performance on a wide range of eHealth tasks. Hence, different types of challenges arise depending on the tool. Specifically, interaction with different eHealth tools and tasks makes different kinds of demands on skills and knowledge. Table 1 [11,15-19] lists some examples of documented skill-related challenges that may lead to barriers to the use of different eHealth tools.

Table 1. Documented skill-related challenges to use of common eHealth tools

eHealth tool	Example of tasks	Examples of skill-related challenges in completing eHealth tasks
Health information portals	Looking up information about treatment options for a health condition	• Identifying appropriate and reliable sources; assessing quality of information • Using effective information retrieval strategies [15] • Understanding complex technical language • Comprehending materials written above recommended reading levels [11]
Personal health records	Entering personal information into medical record	• Having computer skills to effectively use all the different features and tools • Being familiar with health concepts to enter and extract appropriate information in record [16]
Telemedicine or teleconsultation applications	Communicating with health care providers	• Effectively using communication tools • Interpreting and using health information appropriately for self-care activities [17]
Decision-support tools	Evaluating and weighing evidence to inform a decision	• Understanding risk and uncertainty [18] • Obtaining and evaluating evidence-based information
Online support or chat groups	Participating in discussion forum	• Communicating ideas clearly; adhering to online social etiquette and group norms • Effectively sharing information without compromising one's privacy [19]

There is a divide between what consumers can reasonably be expected to do and the demands and available resources of different tools. Various research efforts, in areas such as educational media, health literacy, and numeracy research, have tried to bridge this gulf by addressing user knowledge and competence, and improving resources. Addressing access and skills barriers has helped underserved and vulnerable populations to use technology in terms of managing their health concerns [20,21]. Therefore, it is important to identify barriers and devise solution strategies to eliminate obstacles that reinforce eHealth disparities.

XSL•FO
RenderX

Our approach is an effort to make this a more tractable problem by identifying candidate explanatory constructs and employing cognitive task analysis (CTA) methods to new applications. To the best of our knowledge, this is a unique effort to introduce systematicity to this complex and ill-defined research space. In our view, the success of consumer health informatics initiatives is partially predicated on an understanding of eHealth literacy demands and competencies.

Theoretical Framework

In this research, we endeavored to develop a systematic approach to analyzing competencies across eHealth interventions. The objective of this research was to understand the core skills and knowledge needed to productively use eHealth tools and to develop a set of methods for analyzing eHealth literacy. Previously, we presented a preliminary sketch of our framework for characterizing eHealth literacy task demands [22]. In this study, we explored further application of the framework to characterize human task performance.

The approach draws on two established models: the eHealth literacy model and Bloom's taxonomy of educational objectives. eHealth literacy is defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [23]. The eHealth literacy model describes the set of "fundamental skills consumers require to derive direct benefits from eHealth" [23]. This set of skills establishes an important starting point but does not provide us with a means to discern how different cognitive functions or processes are engaged by different tasks. In addition, eHealth skills may be acquired in different stages, and thus individuals may display different degrees of competence in these skills. We incorporated a second model that is designed to discriminate between various kinds of cognitive processes and describe dimensions of cognitive complexity. Bloom's taxonomy describes the increasing progression in complexity of cognitive aspects of learning, skill acquisition, and performance, and it has been applied to a range of different topic domains [24-26]. Incorporating this model allows us to characterize the central cognitive processes that constitute each literacy type. The eHealth literacy model defines a literacy type or content domain, while the cognitive taxonomy provides a means of realizing how this can be expressed in the context of task performance.

In our framework, we adapted the eHealth literacy model proposed by Norman and Skinner in 2006 as a point of departure. Their model describes six components of eHealth literacy [23]:

- *Computer literacy* describes a wide range of skills from basic knowledge of using a computer, such as opening a browser window, to participating in social networking activities.
- *Information literacy* encompasses the skills to articulate information needs, to locate, evaluate, and use information, and to apply information to create and communicate knowledge [27].
- *Media literacy* is the ability to select, interpret, evaluate, contextualize, and create meaning from resources presented in a variety of visual or audio forms [28].
- *Traditional literacy and numeracy* encompasses reading and understanding written passages, communicating and writing a language coherently, quantitative skills, and the ability to interpret information artifacts such as graphs, scales, and forms [29,30].
- *Science literacy* includes familiarity with basic biological concepts and the scientific method, as well as the ability to understand, evaluate, and interpret health research findings using appropriate scientific reasoning [31].
- *Health literacy* is the acquisition, evaluation, and appropriate application of relevant health information that allows consumers to communicate about health, make health decisions, and use health services [11,32].

Although this model of eHealth literacy is not inclusive of all factors that may influence the use of eHealth (e.g., knowledge of the social and cultural norms involved in participating in a support forum), it is our contention that these six literacy types constitute the set of core skills and knowledge domains.

We selected a second model that explains variation in task performance along an increasing continuum of cognitive demands. Bloom's taxonomy is a well-known taxonomy developed in 1956 and was revised and updated in 2002 [33]. The taxonomy classifies levels of intellectual behavior in learning and has been applied to develop educational objectives and curriculum, assess learning, and create test items [33]. The cumulative hierarchy structure requires achievement of a prior skill before acquiring the next dimension of complexity, but the boundaries between these levels are not rigid. These six cognitive process dimensions are defined [34] as follows:

- *Remembering* is retrieving, recognizing, and recalling relevant knowledge from long-term memory.
- *Understanding* includes constructing meaning from oral, written, and graphic messages through interpreting, classifying, summarizing, inferring, comparing, and explaining.
- *Applying* involves using knowledge to execute a procedure.
- *Analyzing* comprises breaking material into constituent parts, and determining how the parts relate to one another and to an overall structure or purpose.
- *Evaluating* involves making judgments based on criteria and standards.
- *Creating* consists of putting elements together to form a coherent or functional whole in a new pattern or structure.

An overlay of Bloom's taxonomy across the six eHealth literacies provides a framework to characterize and describe the different levels of cognitive demands within each of the six facets of eHealth literacy. It provides a structure to the analysis of human performance on eHealth tasks, allowing a differentiation of cognitive processes as well as of level of knowledge and skill.

The aim of this study was to characterize the constituent elements of eHealth literacy in performing tasks. The hypothesis was that this method can be used to elucidate the barriers to effective task performance.

XSL•FO
RenderX

Methods

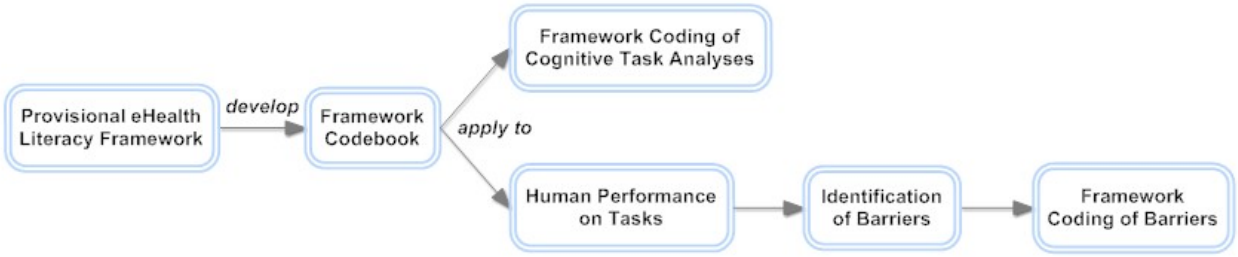
Overview of the Framework and Method

The framework can be expressed as a matrix with the six facets of eHealth literacy along one axis and the six levels of complexity along the other axis, resulting in 36 combined categories. In our framework, we further separated the category of traditional literacy and numeracy into reading, writing, and numeracy and analyzed each separately, as shown in Table 2, such that there are a total of eight different literacy types. In our preliminary application of the framework, it was evident that this revision was necessary to achieve sufficient level of detail for analysis. We defined the criteria for each of the cells through an iterative process of review and adaptation, drawing on evidence from peer-reviewed articles discussing eHealth and each type of literacy. This matrix of eHealth literacy and complexity definitions constituted the framework and codebook, providing the foundation for analysis. The framework coding can be used in two complementary ways. As Figure 1 shows, we employed a CTA and used it to characterize the demands of eHealth tasks with reference to specific tools. We also used the same categorical scheme to describe human performance on these tasks. The basis of the methodological framework involved coordinating task analysis and analysis of human performance.

Table 2. Framework shown as a matrix of literacy types and cognitive complexity levels

Literacy type	Increasing levels of cognitive complexity (Bloom's taxonomy)					
	Remembering	Understanding	Applying	Analyzing	Evaluating	Creating
Computer						
Information						
Media		*The contents of this table are intentionally left blank. This table illustrates the structure of the framework coding tool, which can be used by researchers to map skill demands to the corresponding framework code in each cell of the table.*				
Reading						
Writing						
Numeracy						
Science						
Health						

Figure 1. Process of employing a framework to characterize eHealth demands and barriers.



Application 1: Cognitive Task Analysis

To characterize eHealth literacy demands, we employed CTA, a cognitive engineering method that decomposes a task to uncover knowledge, goal structures, thought processes, and strategies underlying task completion [35,36]. Expert analysts carried out CTA by performing the task themselves, eliciting both information-processing demands of a task and the kinds of domain-specific knowledge required [37]. In the study of health technologies, CTA is most commonly used to study system usability, devise training protocols, or analyze technology-mediated work [38]. We applied CTA in a novel application, to characterize the actions, either behavioral or cognitive, and the knowledge needed to execute an eHealth task. For each task, CTA was used to enumerate the action and knowledge steps used to complete the specified task and to identify the constituent skills required to complete each step. Next, the codebook was used to select the types of literacy that describe the knowledge used in each step. We then determined the kinds of cognitive operations involved in the task that would provide us with a complexity level. For example, a step may require reading a text passage in order to follow the directions in the passage. To apply the framework code, we first identified that this step requires *reading literacy*, and then determined that reading is required at the *applying* level of complexity to use the information in the passage appropriately. The step also required *information literacy* at the *understanding* level of complexity to be able to meet the appropriate information need while reading the passage. Most steps require more than one type of literacy.

In prior work, we illustrated the application of the framework analysis with CTA of three information-seeking tasks [22]. When applied to eHealth tasks, the framework provided

XSL•FO
RenderX

illuminating representations of a task, displaying the configurations of eHealth literacy and cognitive demands for each task. The preliminary findings suggested that the approach enabled deeper exploration of the complex relationships and interactions of the different types of literacy. Our current research explored further applicability of the framework by applying the approach to a new task category, decision making, and to a wider range of health domains. We also applied the method to analysis of human performance and explored how the framework elucidates and diagnoses barriers encountered.

Application 2: Analysis of Human Performance in Task Completion

In the second step of our method, we recruited 20 users to perform the same tasks and observed their performance. These individuals were active computer users but had no previous experience using the website employed in the study, the Consumer Reports Health website (http://www.consumerreports.org/health), a resource that helps consumers make evidence-based decisions related to health issues.

Participants were asked to verbalize their thoughts (a think-aloud protocol) while completing the task and to explain their answers. The think-aloud protocol can reveal any hesitation, confusion, or misunderstanding while completing the task [39]. It can also reveal insights into reasoning and decision-making processes [40]. While the participants were completing the task, guidance was provided when necessary to help them complete a task or to reroute them from a potentially fruitless path. Each session was audio recorded, and we used Morae 3.1 video-analytic software (TechSmith Corporation, Okemos, MI, USA) to capture all actions on the computer screen.

A step-by-step analysis of the participants' performance was done based on the audio recording, video capture, and notes taken during observation of the session. The measures employed were (1) accuracy of response to each question in the task, and (2) any barriers encountered at each step toward completing the task. Task responses were scored according to a scoring scheme comprising specific criteria defining scores of 0 (incorrect), 1 (partially correct), or 2 (correct). In our analysis, barriers were defined as events where participants struggled and may have been unable to make progress in the task or may have required some problem-solving steps before moving forward in the task. Barriers may be indicated when participants required prompts, asked questions, or made errors. A prompt was noted if the researcher provided some verbal assistance to participants, such as directing them to appropriate information or reminding them about the next step of the task. Questions occurred when participants asked a question, expressed confusion, or requested guidance from the researcher. An error was documented if there was a misstep or misinterpretation of information or system response, such as misunderstanding search results. For each barrier event, the framework coding could be applied to categorize the nature of the participant's problem in terms of a type of literacy. For example, difficulty with scrolling would be categorized as difficulty with a computer literacy skill, whereas struggling with text passages would be categorized as difficulty with reading skills. We also matched each event with the corresponding step in the task completion process in which it occurred.

Example of Applying the Framework and Method

We applied these methods to the analysis of a particular task that required a series of information-seeking and decision-making steps. We selected the Consumer Reports Health website because, in our judgment, it is a high-quality and well-designed site that reflects a genuine understanding of consumers' needs. The task question (see Textbox 1) asked users to consider criteria comparing three different hospitals, demonstrate understanding of the information, and interpret the evidence presented.

Textbox 1. Task Question Requiring a Series of Information-Seeking and Decision-Making Steps

In the Doctors & Hospitals page, read the article "How-to guide to choosing a hospital" which can be found at the bottom of the page.

Look up the hospital ratings for **all** hospitals in the New York, NY region.

Next, on the ratings page, *use the Compare feature* to compare New York Presbyterian Hospital, Lenox Hill Hospital, and Bellevue Hospital Center.

A. Identify the hospital that is least aggressive on the "Aggressive or Conservative" scale.

   What do these ratings of "Aggressive or Conservative" tell you about the hospital?

B. Identify the hospital with the highest "Average Cost to Patient".

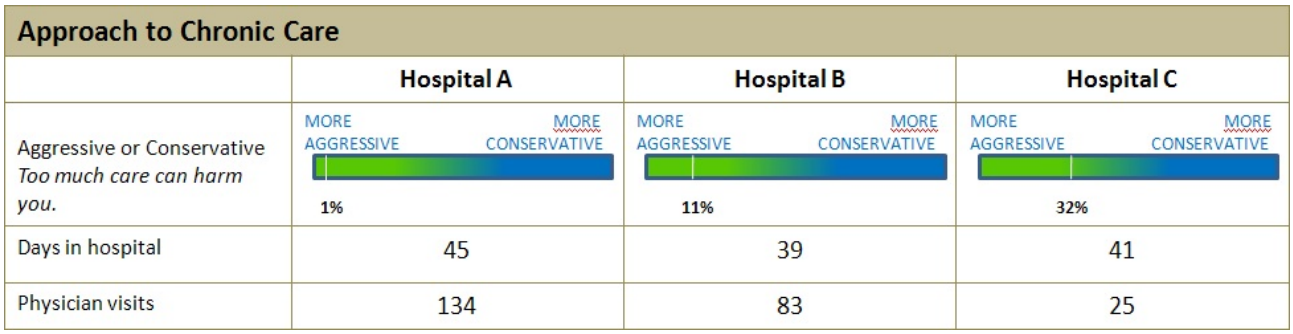
C. Of these 3 hospitals, select the hospital that you would want to go to for a surgical procedure, and discuss what criteria are most important in your decision.

Figure 2 shows a representation of the aggressive/conservative scale needed to interpret the "aggressive or conservative" continuum. These rows were extracted from a table on the pertinent Consumer Reports Health webpage. The aggressive-to-conservative continuum is one way in which the Consumers Union rates its hospitals. Hospitals that keep people with chronic diseases hospitalized for more days during the last 2 years of their lives are rated as aggressive. Hospitals that provide the least amount of doctor's visits and shorted hospitalizations in those final years of life are considered conservative. We used example tasks to explore the reliability of the framework coding scheme. Two different raters used the codebook to classify task demands on two different tasks: an information-seeking and a decision-making task. The raters later used the codebook to also classify the barriers encountered by a subset of three different participants. For each type of coding, interrater reliability was assessed on two different dimensions of the coding: (1) type of literacy, using Cohen's

XSL•FO
RenderX

kappa, and (2) level of cognitive complexity, using Spearman correlation coefficient. The assignment of a cognitive complexity code cannot be coded independent of literacy type. Therefore, cognitive complexity was calculated on the subset of codes in which both raters reached agreement on the literacy code.

Figure 2. Representation of the aggressive/conservative scale. The rows are as they appear in the actual table on the webpage except for the top row, which is included for clarity.



Results

To illustrate the application of the framework and method, we present the results from the analysis of the task question asking about hospital ratings.

Application 1: Cognitive Task Analysis

Interrater reliability was calculated for coding of the CTA. Cohen's kappa for literacy was .91 and Spearman correlation coefficient for cognitive complexity was .92, suggesting high levels of agreement for both dimensions. Table 3 shows the CTA of an excerpt of this task (steps 10–16 of the entire task), from the steps for selecting the three specific hospitals for comparison to interpreting the aggressive or conservative scale.

Table 3. Application of the framework coding to steps 10–16 of the task

Step	Skills and knowledge required to complete step	Framework code from CTA[a]
10	Recognize the results page as a table of hospitals and their ratings. Scroll to see whole table.	Computer 3, information 4, numeracy 4, reading 1
11	Recognize the “compare” feature, and that checkboxes for the desired hospitals are required to use this feature. Select the appropriate checkboxes for the three hospitals.	Computer 3, information 3, reading 2
12	Recognize results as a table of the three selected hospitals with their detailed ratings. Scroll to see whole table.	Computer 3, information 4, numeracy 4, reading 2
13	Scroll to locate the “aggressive or conservative” row in the table. Interpret and understand the labels for the aggressive/conservative scale.	Computer 3, information 4, numeracy 4, reading 2
14	Identify the least aggressive rating and answer the information need.	Information 5, numeracy 4, reading 2, writing 2
15	Click on the “learn more” link. Find the newly opened window. Scroll down to find the text about aggressive/conservative hospitals. Read and understand text.	Computer 3, information 4, health 4, reading 3
16	Articulate understanding of what aggressive/conservative means.	Health 4, writing 3

[a] Cognitive task analysis, by increasing complexity: 1 = remembering, 2 = understanding, 3 = applying, 4 = analyzing, 5 = evaluating, 6 = creating.

Completing these series of steps required the participant to navigate to the table, locate the relevant information, and interpret the data in the table. The aggressive/conservative scale (see Figure 2) corresponds to step 13 in Table 3. Each step was coded with the corresponding framework codes that describe the eHealth literacy and complexity level used to complete that step. For example, step 10 required a combination of four types of eHealth literacy: (1) information literacy at the analyzing level of complexity (information 4) was required to interpret and evaluate the results page, (2) numeracy at the analyzing level of complexity (numeracy 4) was required to interpret the results table, (3) computer literacy at the applying level of complexity (computer 3) was required to navigate and interact with the table, and (4) reading was required at the remembering level (reading 1) to make sense of the information in the table. The steps required different combinations of literacy types, ranging from a combination of two to four types of literacy. The highest complexity level of any eHealth literacy required was level 5, evaluating. Reading and information literacy were required for most of the steps in this excerpt and appeared more frequently than the other literacies.

Table 4 summarizes the results of coding the CTA for the entire task. For the whole task, reading was used most often, in 18 of the 20 steps (90%). Information literacy (17 of 20 steps) and computer literacy (15 of 20 steps) were also used often. The frequent use of these skills suggests that they are essential to

XSL•FO
RenderX

completing the task and are useful skills to promote among health care consumers. Media literacy was not required for any steps in this task given that the website was already selected as part of the task. Information literacy was required at level 5 (evaluating) for two of three questions, suggesting that high levels of information literacy are necessary to be able to carry out all components of this task. As this task was primarily an information-seeking task, it is not surprising that information literacy was required frequently and at high levels of cognitive demand. Numeracy was required at level 4 (analyzing) for all three questions, to understand the information in different representational formats and to interpret the data in the table of hospital ratings. Question C was the only question to require science literacy, which was used to weigh evidence in making a decision about selecting a hospital based on the criteria presented. Question C also required the most skills at level 5 (evaluating), for two different types of literacy, suggesting that it had the highest complexity demands across the whole task. Question B was the only question to require any skills at level 2 (understanding); Question B had the lowest complexity demands relative to the other questions. The highest complexity level required across the whole task was level 5 (evaluating).

Table 4. Summary of task demands from cognitive task analysis

Literacy type	Question A	Question B	Question C	Whole task
Media	0%[a]	0%	0%	0%
	No complexity	No complexity	No complexity	No complexity
Computer	50%	50%	50%	75%
	Applying (3)	Applying (3)	Applying (3)	Applying (3)
Health	50%	0%	100%	35%
	Analyzing (4)	No complexity	Analyzing (4)	Analyzing (4)
Information	75%	100%	50%	85%
	Analyzing (4)	Evaluating (5)	Evaluating (5)	Evaluating (5)
Reading	75%	100%	50%	90%
	Applying (3)	Understanding (2)	Applying (3)	Applying (3)
Writing	50%	50%	50%	20%
	Analyzing (4)	Understanding (2)	Evaluating (5)	Evaluating (5)
Numeracy	50%	100%	50%	30%
	Analyzing (4)	Analyzing (4)	Analyzing (4)	Analyzing (4)
Science	0%	0%	100%	10%
	No complexity	No complexity	Applying (3)	Applying (3)
Total number of steps	4	2	2	20[b]

[a] For the task, the following is displayed: the proportion (percentage) of steps that use that eHealth literacy and the highest level of cognitive complexity used in that literacy (number and complexity level).

[b] Total number of steps for whole task includes a series of 12 navigational steps leading up to questions A, B, and C.

Application 2: Analysis of Human Performance in Task Completion

The framework coding was then applied to the task performance. Interrater reliability was calculated for the coding of task performance. Spearman correlation coefficient for cognitive complexity was .88, suggesting high agreement. Cohen's kappa for literacy was .68, suggesting lower but sufficient agreement to meet the minimum standard. The results from a single user are displayed in Table 5.

XSL•FO
RenderX

Table 5. Mapping the framework coding for steps 10–16 to a participant's performance on the task

Step	Skills and knowledge required to complete step	Framework code from CTA[a]	Events that indicate barriers	Framework code for barrier
10	Recognize the results page as a table of hospitals and their ratings. Scroll to see whole table.	Computer 3, information 4, numeracy 4, reading 1	Participant asks: "Aggressive or conservative scale—where's that?" Participant is not on the correct page yet, needs to navigate to the next page first.	Computer 2, information 2
11	Recognize the "compare" feature, and that checkboxes for the desired hospitals are required to use this feature. Select the appropriate checkboxes for the three hospitals.	Computer 3, information 3, reading 2	Researcher prompts: "Use the 'compare' feature."	Computer 3, information 1
			Error: participant clicks on "compare" without having selected the hospitals to compare.	Computer 3
			Researcher prompts: "In order to compare the three, you want to select all three together."	Computer 3, information 2
12	Recognize results as a table of the three selected hospitals with their detailed ratings. Scroll to see whole table.	Computer 3, information 4, numeracy 4, reading 2	No barrier encountered during this step.	None
13	Scroll to locate the "aggressive or conservative" row in the table. Interpret and understand the labels for the aggressive/conservative scale.	Computer 3, information 4, numeracy 4, reading 2	Participant confused by the multiple parts of the task question. Researcher prompts: "Look at this part of the question first."	Information 1
			Participant scrolls up and down, and finds the aggressive/conservative scale. Starts to read ahead to the next question. Researcher prompts again: "Try this question first—the hospital that is least aggressive."	Information 1
			Participant asks: "Where does it tell you which is least or most aggressive/conservative? In this area here?" (pointing to the scale).	Information 2, numeracy 4
14	Identify the least aggressive rating and answer the information need.	Information 5, numeracy 4, reading 2, writing 2	Participant stares at scale, confused. Researcher prompts: "What do you think the scale is telling you; how are you reading the scale?"	Numeracy 4
			Participant is very confused by the scale, and answers: "The one that is more conservative is 32%, Bellevue. Least aggressive, Lenox Hill? I'm trying to understand this." (incorrect)	Numeracy 4
15	Click on the "learn more" link. Find the newly opened window. Scroll down to find the text about aggressive/conservative hospitals. Read and understand text.	Computer 3, information 4, health 4, reading 3	Participant is unsure how to approach the next question. Researcher rewords the question and explains what the question is asking.	Information 1, reading 2, information 2, reading 2
			Participant clicks on the "learn more" link and scrolls down the page, but cannot find the relevant text. Participant scrolls past the relevant passage. Researcher prompts: "You just missed the description on the page."	
16	Articulate understanding of what aggressive/conservative means.	Health 4, writing 3	Participant reads the text passage, then answers: "More doctors visit overall for aggressive/conservative care...fewer days in the hospital." (incorrect)	Health 3

[a] Cognitive task analysis, by increasing complexity: 1 = remembering, 2 = understanding, 3 = applying, 4 = analyzing, 5 = evaluating, 6 = creating.

This participant scored low on this task, earning 2 out of a total of 6 possible points. The participant encountered 18 barriers while completing this task. In step 10, the participant was looking for a specific piece of information but was on the wrong page; this barrier can be attributed to problems or deficiencies associated with information and computer literacies. In step 14, the participant was confused by the scale and provided an incorrect answer due to misinterpretation of the information presented in the aggressive/conservative scale. This barrier reflects a struggle with numeracy because the participant demonstrated an understanding of numbers as evidenced by the ability to draw inferences about the scale, but was unable to *apply* the knowledge and *analyze* it in different representational formats. Then in step 16, the participant provided an incorrect

XSL•FO
RenderX

answer. The participant was unable to read, interpret, and analyze the health text to extract an accurate description of the terms aggressive and conservative as used in this context; this barrier reflects a struggle with health literacy. The participant required several reminders or explanations of task questions, in steps 11, 13, and 15. These reminders and explanations indicated information literacy barriers, reflecting a lack of recognition and understanding of the nature of the information need.

Summary Results From 20 Participants

A summary of 20 users' task performance results are presented to illustrate the aggregate measures obtained and potential analyses that can be performed using our approach. The users were recruited from the Union Settlement Association and the Columbia Community Partnership for Health Center in New York, NY.

Participants recruited were adults between 18 and 65 years of age; all had basic proficiency with computers and the Internet. A total 14 of 20 (70%) of participants were female, most reported annual incomes below US $30,000, and a majority of participants reported their race as African American or Hispanic. Participants had a range of education backgrounds, with 7 participants reporting high school education, 7 having a college degree, and 6 with a graduate degree.

As Table 4 describes, Question B had the lowest literacy and complexity demands relative to the other questions. Figure 3 shows that participants scored highest on this question, with 16 out of 20 correct answers. Participants struggled most with question A, with only 2 correct answers, and 11 partial answers. Although question C had the highest complexity levels of cognitive demands, 10 out of 20 participants (50%) answered this question correctly. Each question varied in terms of domain knowledge, complexity, and types of demands. Scores merely provide a snapshot of user task performance. Although we can use the scores to compare and contrast task performance across the different task questions, analysis of the barriers impeding task performance can yield additional insight into the resulting participant scores.

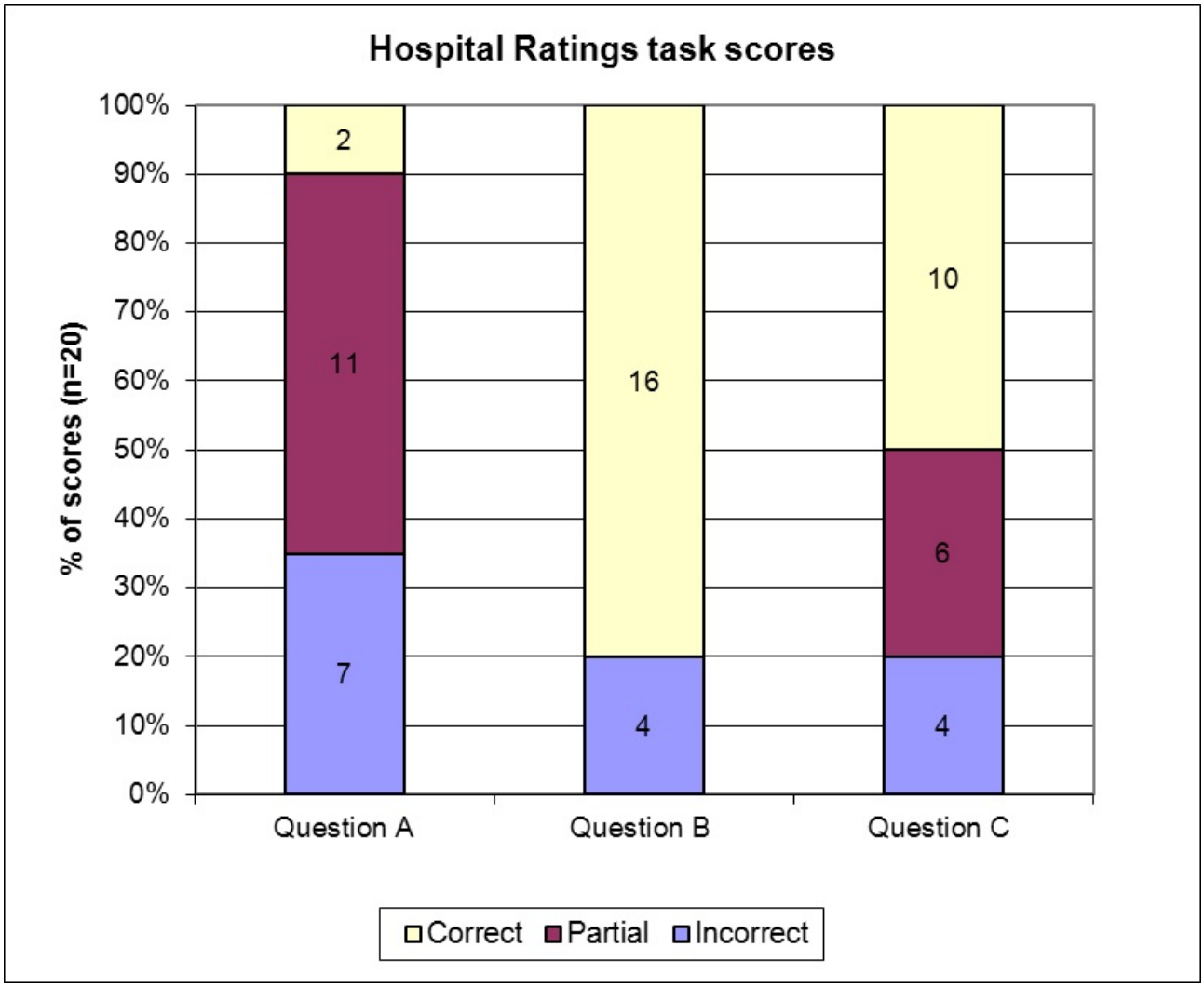
Figure 4 shows the number of barriers encountered by all participants for each step. The most barriers were encountered in step 11, with a total of 51 barriers encountered. This step required users to make the appropriate selections in order to compare the different hospitals selected. Most of the barriers on this step stemmed from unfamiliarity with making the appropriate selections using checkboxes, reflecting inadequate computer literacy. Users encountered a high number of barriers at steps 13 and 15 as well. These steps are both constituents of question A, on which participants scored the lowest of the three questions. This aggregate analysis revealed the steps in which users experienced the most difficulty and exemplifies the patterns of barriers encountered in carrying out those problem steps.

We aggregated the types of barriers encountered by users in a manner similar to the analysis in Clark et al [41], which provided cumulative descriptors of the component barriers encountered across a set of steps and tasks. Figure 5 presents the classifications of literacy type and cognitive demands of barriers encountered in task performance. The same excerpt of steps (steps 10–16) was depicted as in Table 3. Most of the barrier classifications in these steps are due to barriers with information and computer literacy. Step 13, which required understanding question A, caused many barriers at levels 1 (remembering) and 2 (understanding) within information literacy. These barriers primarily involved struggling to identify and interpret the information need. Step 14, which required locating and interpreting the aggressive/conservative scale, led to many numeracy level 4 barriers (analyzing based on representation). Step 15 asked users to describe the meaning of aggressive/conservative in the context of hospitals and health care, and users struggled with finding resources to meet this information need. These barriers are reflected by the majority of barriers being information literacy and computer literacy barriers. Step 16 reflected many health literacy as well as some writing barriers; users struggled with understanding, interpreting, and articulating aggressive/conservative in their own words. The majority of barriers fell in the lower ranges of cognitive demands (levels 1–4). The task demands also required mainly literacies at these lower levels. The patterns of barrier types as revealed by the coding reflected the nature of the task demands and provided insight into the types of barriers that participants encountered.

Overall, within the hospital ratings task, users scored highest on question B and encountered the most barriers in question A. The barriers identified reflected that users struggled primarily with information literacy, computer literacy, and numeracy skills in answering the question and completing the tasks.

XSL•FO
RenderX

Figure 3. Hospital ratings task: distribution of participants' scores on questions A, B, and C, and average (Avg) scores for each question.



XSL•FO
RenderX

Figure 4. Number of barriers encountered by participants in each step, with labels for the steps that constitute questions A, B, and C.

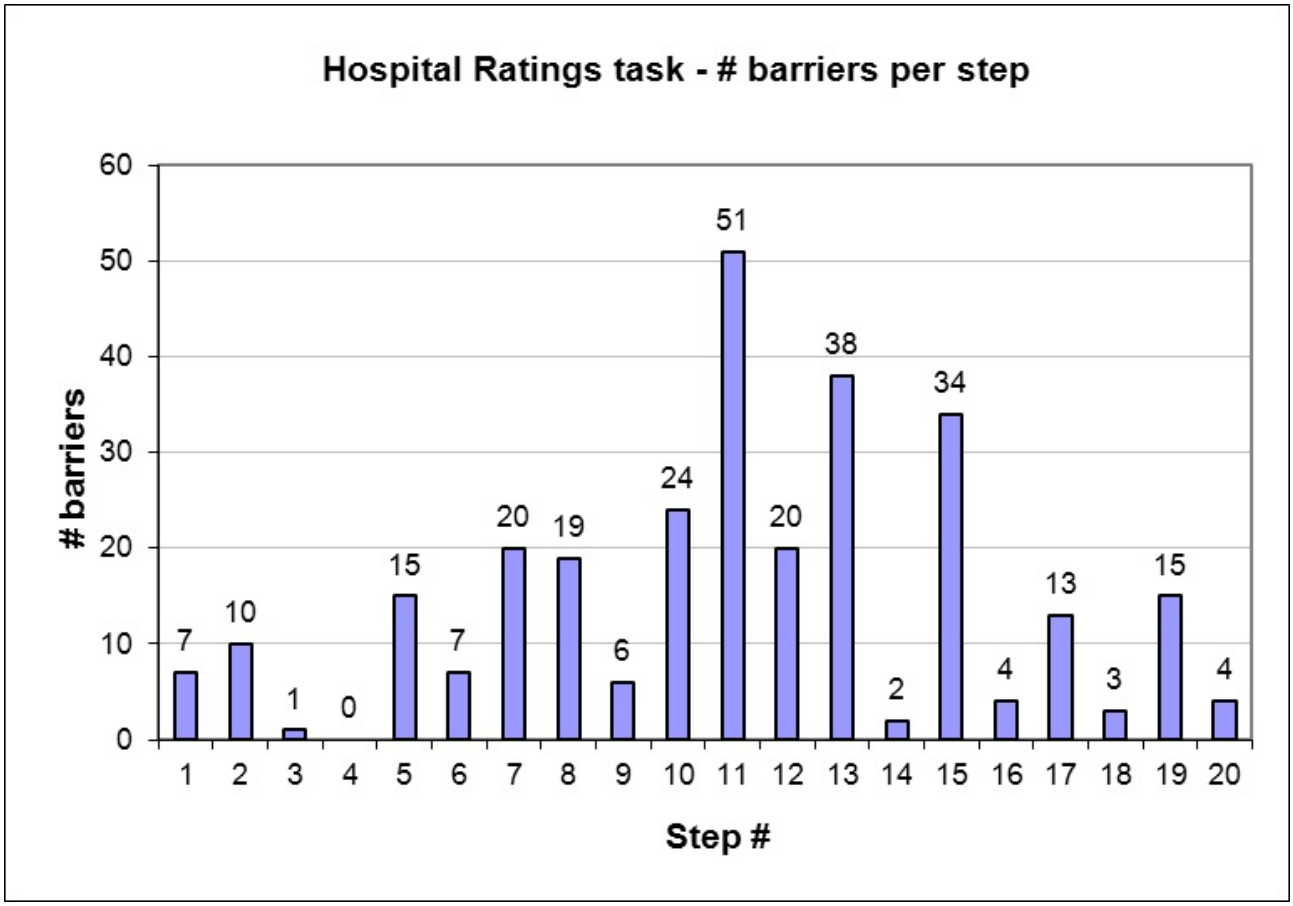
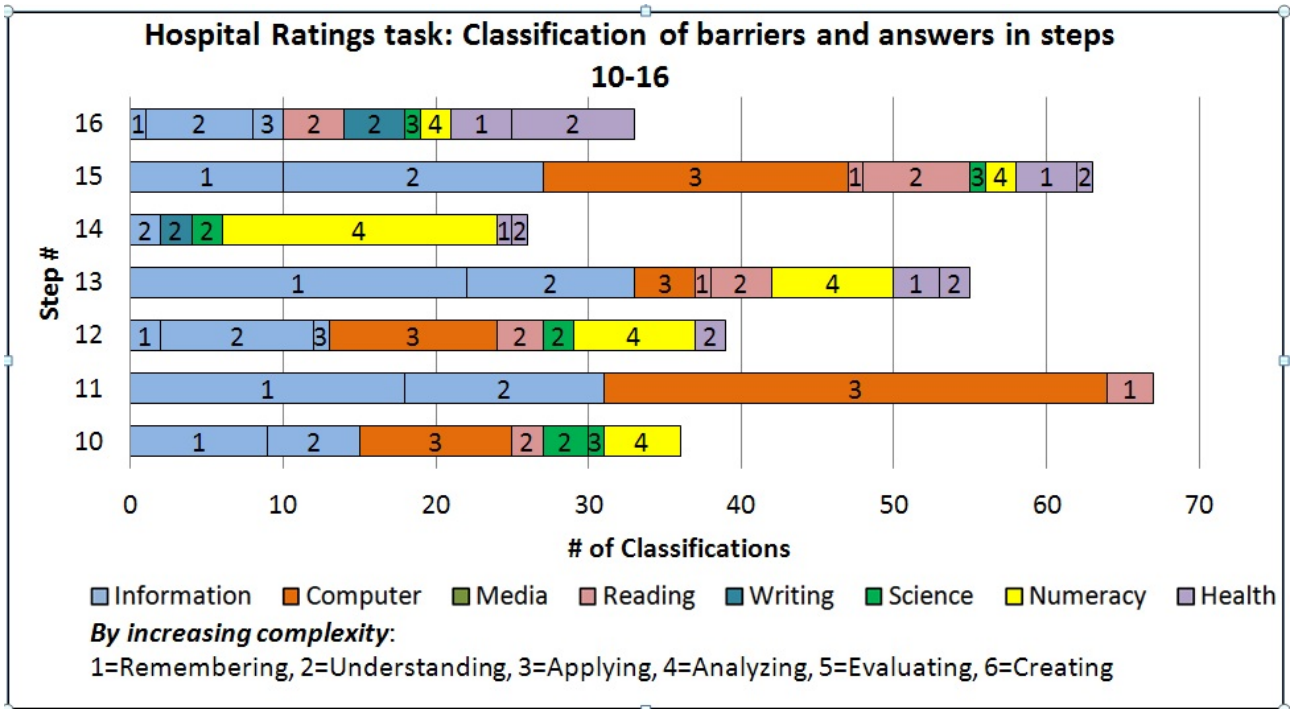


Figure 5. Barriers encountered by participants (n = 20) in steps 10–16, categorized by literacy (color in legend) and complexity level (number in the graph).



Discussion

In this research, we adapted and integrated two existing theoretical models relevant to the analysis of eHealth literacy into a single framework to systematically categorize and describe task demands and user performance on tasks needed by health care consumers in the information age. The method derived from the framework is applied to (1) code task demands

XSL•FO
RenderX

using a CTA, and (2) code user performance on tasks. The analysis shows that the framework can be used to classify task demands as well as the barriers encountered in user performance of the tasks. Our approach can be used to (1) characterize the challenges confronted by participants in performing the tasks, (2) determine the extent to which application of the framework to the CTA can predict and explain the problems encountered by participants, and (3) inform revisions to the framework to increase accuracy of predictions. In this study, we used the methods to document a range of literacy-related barriers that affected performance on eHealth tasks.

The study found that 20 participants experienced some difficulty completing most tasks on a website designed for consumers without some assistance. The most frequent barriers encountered by our sample were challenges with information literacy and computer literacy skills. Specific examples of frequent barriers encountered are struggling with the ability to understand and successfully act on information needs, to interpret a graphical representation of a severity scale, and to effectively use checkboxes to make selections. Conversely, some activities in which we had predicted barriers were discovered to be easier than anticipated. Evaluating health information to inform decisions can be complex and challenging, but users scored well on the question with a decision point.

There is little existing research that systematically analyzes the combined set of eHealth skills needed to attain proficient performance. Other investigators have expanded the scope of health literacy to describe the combinations of skills needed to interact effectively with health information [42] but did not consider technology-related skills, such as computer literacy, that are a core part of eHealth literacy. Our results largely echo findings in prior health numeracy research that users often struggled with interpreting graphical representations of numerical information, which may constitute significant consumer barriers [29]. Our findings also support recommendations to develop tools that aid health care consumers in understanding complex health concepts and to use the information to inform a decision [43]. Usability studies take a similar approach in breaking down task demands to analyze user task performance. Our method is consistent with usability findings that a granular approach to task analysis is essential to reveal potential barriers and inform design improvements, particularly for novice users [44].

Limitations

We view the framework as provisional and subject to more comprehensive validation and elaboration. This will necessitate a larger-scale study with a greater sample size, a more diverse population, and a wider range of tasks. In addition, the participants in the study were not familiar with the Consumer Reports Health website and this may have influenced our findings. Familiarity with content, style, and affordances common to this site would have likely reduced some of the barriers that participants experienced. Further studies should include participants with varying degrees of experience with a particular website or technology.

The analyses in this paper focused on user competencies and did not take into consideration a range of issues, such as usability, or affordances and resources available within specific technology tools. In addition, the methods employed did not take into account individual motivation or attitudes toward technology. Similarly, this cognitive rational framework does not capture emotional and social factors that also play a significant role in decision making. It is well known that health literacy is a major public health issue in the United States affecting a substantial segment of the population [11]. In general, a multitude of environmental and societal factors, such as differential access to the eHealth tools, influence the productive use of technology in health-related contexts. Although these individual and social factors significantly influence task performance, our leading-edge hypothesis is that eHealth literacy is a distinct construct and an important one in consumer health informatics.

As previously described in Table 1, there are many different types of eHealth tools and eHealth tasks. The framework was illustrated using an example task on the Consumer Reports Health website. This website aims to present information simply and comparatively. Consumer Reports has been presenting unbiased and evidence-based comparisons in print form for many years. However, evidence in health is often complex and there may be alternative ways for rendering such information as comprehensible to individuals lower in eHealth literacy. The effective presentation of health evidence is a challenge that continues to plague most health communication and decision aid materials [45]. The website selection was sufficient for the purpose of illustrating the framework. It should be noted that the aggressive/conservative continuum scale is no longer used on the Consumer Union's health site. Further exploration will apply the framework to a wider array of tasks, tools, and health domains.

Further Development of the Framework and Analytic Method

Further studies are needed to determine whether the types of literacy described in this paper sufficiently cover the range of knowledge types that characterize eHealth competency. In addition, although Bloom's taxonomy has an established history of characterizing cognitive dimensions of tasks in educational contexts, we cannot presuppose that the gradations of complexity will seamlessly transfer to eHealth. The results of this analysis suggest that it can be used meaningfully to differentiate and categorize cognitive demands for different literacy skills and can be used to approximate complexity in a range of eHealth tasks.

As discussed, the tasks used in the study did not delve deeply into media, science, and to some extent health literacy. As health consumers choose what resources to use, media literacy will loom large. We anticipate that our methods will be adequate to model the skills and knowledge needed to demonstrate media literacy competency. The problems associated with low health literacy are well documented [11]. Science literacy is a multifaceted construct, and there is ample evidence to suggest that problems associated with science literacy are equally profound. The general public in the United States and other countries have an impoverished understanding of science [46]. Norman and Skinner [23] situate scientific literacy in a broader

XSL•FO
RenderX

context, defining it as "understanding of the nature, aims, methods, application, limitations, and politics of creating knowledge in a systematic manner." The framework employs a CTA approach that places a strong emphasis on skills and action. This may not capture other dimensions of science literacy such as understanding biological mechanisms of disease and critical appraisal of the scientific process. These aspects come into play in situations such as when an individual must understand the consequences of a therapeutic regimen or decide whether to enroll in a randomized controlled clinical trial. Clearly, we would need a broader array of concepts and a richer set of representations than those offered by the CTA stepwise analytic method to model such knowledge and causal inferences associated with its application in the context of health.

The proposed framework provides a basis for the development of an eHealth competence model. Such a model would yield insight into the specific skills and knowledge needed to perform at a proficient or higher level on system-specific instances of eHealth tasks, such as seeking information about hypertensive therapies on the WebMD site. The current set of framework-based methodological tools lends greater utility to the consumer health research community than to communities of practitioners and designers. Applying this method is time intensive and requires moderate expertise in the areas of cognition and human–computer interaction. We anticipate that the framework would give rise to simpler, more specific instruments (for example, in the form of a set of questions or heuristics) that could measure eHealth demands for a particular task and population as realized in a particular system or device. An analogy would be Nielsen's heuristic evaluation method [47], which has made it possible for teams of developers to conduct basic usability evaluations without extensive training or prohibitive time commitments.

With further investigation, we envision that the framework and analytic approach can be a potentially powerful generative research tool for development of design guidelines of computer-based tools, evaluation heuristics, task-based eHealth literacy assessment, and educational objectives to increase consumer eHealth skills. For example, the framework could form the basis for development of a matching algorithm to identify appropriate tools for users with different skill sets. In particular, this framework and analysis method can be used with health care consumers with low eHealth skills to better understand barriers and to develop educational media or other mediating tools to facilitate engagement with and benefit from eHealth. Barriers fall on a continuum ranging from routine abilities (recognizing how to use widgets) to complex conceptual challenges (deriving inferences from health text). The proposed framework systematically characterizes eHealth barriers, which in turn enables more precise definition within the solution space of methods to overcome those barriers.

Conclusions

In our view, this framework provides a systematic and potentially rigorous approach for analyzing eHealth competencies, which is a challenge of considerable complexity and great significance. Advances in technologies, such as Web 2.0 and social networking functionalities, offer new and ever-changing modes for consumers to interact with and manage health information. In the current environment where eHealth interventions are being developed without a thorough understanding of the consumers, efforts, and resources can be better focused to improve adoption and use rates as well as benefit from use. Unfortunately, these barriers disproportionately affect those who are most vulnerable and may actually serve to exacerbate disparities rather than bridge them. There is no doubt that consumers will be expected to assume a greater role in their health management in coming years, and low eHealth literacy will continue to be a barrier to productive participation. Progress in eHealth research will be integral to the success of consumer health applications and for reducing barriers to the use of those applications.

Acknowledgments

This work is supported by grants from the National Institute for Nursing Research (1R21NR010710) and the National Library of Medicine (1R21LM01068801) awarded to David Kaufman. We thank Suzanne Bakken, Jessica Ancker, Sara Czaja, Noemie Elhadad, Lisa Matthews and Alisabeth Shine for their help in preparing and reviewing this manuscript. We also thank David Nocenti and Union Settlement Association, Bernadette Boden-Albala and the Columbia Community Partnership for Health in recruiting participants. We appreciate Beth Nash and Consumers Union's support in gaining access to their site. We thank the Department of Biomedical Informatics at Columbia University for their support. Connie Chan was supported by NLM pre-doctoral fellowship T 15-LM007079.

Conflicts of Interest

None declared

Authors' Contributions

This work originated with CC's dissertation work. DK was her thesis adviser. All facets of this work from its inception to the completion of this manuscript represent a close collaboration.

References

1. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]

XSL•FO
RenderX

2. Jimison H, Gorman P, Woods S, Nygren P, Walker M, Norris S, et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep Technol Assess (Full Rep) 2008 Nov(175):1-1422. [Medline: 19408968]
3. Demiris G, Afrin LB, Speedie S, Courtney KL, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]
4. Karkalis GI, Koutsouris DD. E-health and the Web 2.0. In: Proceedings of the ITAB. 2006 Presented at: The International Special Topic Conference on Information Technology in Biomedicine; Oct 26-28, 2006; Ioannina, Greece.
5. Giustini D. How Web 2.0 is changing medicine. BMJ 2006 Dec 23;333(7582):1283-1284. [doi: 10.1136/bmj.39062.555405.80] [Medline: 17185707]
6. Viswanath K, Kreuter MW. Health disparities, communication inequalities, and eHealth. Am J Prev Med 2007 May;32(5 Suppl):S131-S133. [doi: 10.1016/j.amepre.2007.02.012] [Medline: 17466818]
7. Fox S. The engaged e-patient population. Washington, DC: Pew Internet & American Life Project; 2008 Aug 26. URL: http://www.pewinternet.org/Reports/2008/The-Engaged-Epatient-Population/The-Engaged-E-patient-Population.aspx [accessed 2011-01-29] [WebCite Cache ID 5w6KwYvyD]
8. Chapman G, Elstein A. Cognitive processes and biases in medical decision making. In: Chapman G, Sonnenberg F, editors. Decision Making in Health Care: Theory, Psychology, and Applications. New York, NY: Cambridge University Press; 2000:183.
9. 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]
10. Jaffery JB, Becker BN. Evaluation of eHealth web sites for patients with chronic kidney disease. Am J Kidney Dis 2004 Jul;44(1):71-76. [Medline: 15211440]
11. McCray AT. Promoting health literacy. J Am Med Inform Assoc 2005;12(2):152-163 [FREE Full text] [doi: 10.1197/jamia.M1687] [Medline: 15561782]
12. McNeill LH, Puleo E, Bennett GG, Emmons KM. Exploring social contextual correlates of computer ownership and frequency of use among urban, low-income, public housing adult residents. J Med Internet Res 2007;9(4):e35 [FREE Full text] [doi: 10.2196/jmir.9.4.e35] [Medline: 18093903]
13. Cashen MS, Dykes P, Gerber B. eHealth technology and Internet resources: barriers for vulnerable populations. J Cardiovasc Nurs 2004;19(3):209-14; quiz 215. [Medline: 15191264]
14. Fox S. Digital divisions. Washington, DC: Pew Internet & American Life Project; 2005 Oct 05. URL: http://www.pewinternet.org/~/media//Files/Reports/2005/PIP_Digital_Divisions_Oct_5_2005.pdf.pdf [accessed 2011-09-22] [WebCite Cache ID 61tPo1406]
15. Zeng QT, Kogan S, Plovnick RM, Crowell J, Lacroix EM, Greenes RA. Positive attitudes and failed queries: an exploration of the conundrums of consumer health information retrieval. Int J Med Inform 2004 Feb;73(1):45-55. [doi: 10.1016/j.ijmedinf.2003.12.015] [Medline: 15036078]
16. Kim EH, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 2009;11(4):e44 [FREE Full text] [doi: 10.2196/jmir.1256] [Medline: 19861298]
17. Kaufman DR, Patel VL, Hilliman C, Morin PC, Pevzner J, Weinstock RS, et al. Usability in the real world: assessing medical information technologies in patients' homes. J Biomed Inform 2003 Apr;36(1-2):45-60. [Medline: 14552846]
18. McCaffery KJ, Smith SK, Wolf M. The challenge of shared decision making among patients with lower literacy: a framework for research and development. Med Decis Making 2010;30(1):35-44. [doi: 10.1177/0272989X09342279] [Medline: 19692709]
19. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23. [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
20. Eng TR, Maxfield A, Patrick K, Deering MJ, Ratzan SC, Gustafson DH. Access to health information and support: a public highway or a private road? JAMA 1998 Oct 21;280(15):1371-1375. [Medline: 9794322]
21. Botts N, Horan TA. Bridging care communication and health management within diverse and underserved populations. In: Proceedings. 2008 Presented at: 14th Americas Conference on Information Systems; Aug 14-17, 2008; Toronto, ON, Canada.
22. Chan CV, Matthews LA, Kaufman DR. A taxonomy characterizing complexity of consumer eHealth Literacy. AMIA Annu Symp Proc 2009;2009:86-90. [Medline: 20351828]
23. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
24. Larkin BG, Burton KJ. Evaluating a case study using Bloom's Taxonomy of Education. AORN J 2008 Sep;88(3):390-402. [doi: 10.1016/j.aorn.2008.04.020] [Medline: 18790101]
25. Pickard M. The new Bloom's taxonomy: an overview for family and consumer sciences. J Fam Consum Sci Educ 2007;25(1):45-55.

XSL•FO
RenderX

26. Reeves M. An application of Bloom's taxonomy to the teaching of business ethics. J Bus Ethics 1990;9(7):609-616. [doi: 10.1007/BF00383217]
27. Catts R, Lau J. UNESCO. 2008. Towards Information Literacy Indicators URL: http://www.uis.unesco.org/template/pdf/cscl/InfoLit.pdf [accessed 2011-01-29] [WebCite Cache ID 5w6Keaj6o]
28. Thoman E. Skills and strategies for media education. Educ Leadership 1999;56(5):50-54.
29. Ancker JS, Kaufman D. Rethinking health numeracy: a multidisciplinary literature review. J Am Med Inform Assoc 2007;14(6):713-721 [FREE Full text] [doi: 10.1197/jamia.M2464] [Medline: 17712082]
30. Rudd RE, Moeykens BA, Colton TC. Health and literacy: a review of medical and public health literature. In: Comings J, Garners B, Smith C, editors. Annual Review of Adult Learning and Literacy. San Francisco, CA: Jossey-Bass; 2000:158-199.
31. Laugksch RC. Scientific literacy: a conceptual overview. Sci Educ 2000;84(1):71-94. [doi: 10.1002/(SICI)1098-237X(200001)84:1<71::AID-SCE6>3.0.CO;2-C]
32. Rudd R, Kirsch I, Yamamoto K. Policy Information Center, Educational Testing Service. 2004. Literacy and Health in America: Policy Information Report URL: http://www.ets.org/Media/Research/pdf/PICHEATH.pdf [accessed 2011-01-29] [WebCite Cache ID 5w6NGxeNj]
33. Krathwohl DR. A revision of Bloom's taxonomy: an overview. Theory Into Pract 2002;41(4):212-218.
34. Amer A. Reflections on Bloom's revised taxonomy. Electronic J Res Educ Psychol 2006(1):213-230 [FREE Full text] [WebCite Cache ID 5z6HDGJrQ]
35. Roth E, Patterson E, Mumaw RMarciniak JJ. Cognitive engineering: issues in user-centered system design. In: Encyclopedia of Software Engineering. 2nd edition. New York, NY: John Wiley; 2002:163-179.
36. Patel VL, Arocha JF, Kaufman DR. A primer on aspects of cognition for medical informatics. J Am Med Inform Assoc 2001;8(4):324-343 [FREE Full text] [Medline: 11418539]
37. Patel V, Kaufman D. Cognitive science and biomedical informatics. In: Shortliffe E, Cimino J, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. 3rd edition. New York, NY: Springer; 2006:133-185.
38. Schraagen J, Chipman S, Shalin V. Cognitive Task Analysis. Mahwah, NJ: Erlbaum; 2000.
39. Ericsson KA, Simon HA. Protocol Analysis: Verbal Reports as Data. Revised edition. Cambridge, MA: MIT Press; 1993.
40. Cotton D, Gresty K. Reflecting on the think-aloud method for evaluating e-learning. Br J Educ Technol 2006;37(1):45-54. [doi: 10.1111/j.1467-8535.2005.00521.x]
41. Clark MC, Czaja SJ, Weber RA. Older adults and daily living task profiles. Hum Factors 1990 Oct;32(5):537-549. [Medline: 2074108]
42. Zarcadoolas C, Pleasant A, Greer DS. Understanding health literacy: an expanded model. Health Promot Int 2005 Jun;20(2):195-203 [FREE Full text] [doi: 10.1093/heapro/dah609] [Medline: 15788526]
43. Nielsen-Bohlman L, Panzer AM, Kindig DA. Health Literacy: A Prescription to End Confusion. Washington, DC: National Academies Press; 2004.
44. Britto MT, Jimison HB, Munafo JK, Wissman J, Rogers ML, Hersh W. Usability testing finds problems for novice users of pediatric portals. J Am Med Inform Assoc 2009;16(5):660-669 [FREE Full text] [doi: 10.1197/jamia.M3154] [Medline: 19567793]
45. Epstein RM, Alper BS, Quill TE. Communicating evidence for participatory decision making. JAMA 2004 May 19;291(19):2359-2366 [FREE Full text] [doi: 10.1001/jama.291.19.2359] [Medline: 15150208]
46. National Science Board. Science and Engineering Indicators 2006. Arlington, VA: National Science Foundation; 2006. URL: http://www.nsf.gov/statistics/seind06/pdf/volume1.pdf [accessed 2011-05-31] [WebCite Cache ID 5z6I0m8Xy]
47. Nielsen J. Heuristic evaluation. In: Nielsen J, Mack RL, editors. Usability Inspection Methods. New York, NY: Wiley; 1994:25-62.

Edited by G Eysenbach; submitted 02.02.11; peer-reviewed by C Norman, P Dykes; comments to author 25.04.11; revised version received 05.07.11; accepted 06.07.11; published 17.11.11.

Please cite as:
Chan CV, Kaufman DR
A Framework for Characterizing eHealth Literacy Demands and Barriers
J Med Internet Res 2011;13(4):e94
URL: http://www.jmir.org/2011/4/e94/
doi:10.2196/jmir.1750
PMID:22094891

©Connie V Chan, David R Kaufman. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 17.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

eHealth Literacy Among College Students: A Systematic Review With Implications for eHealth Education

Michael Stellefson[1], PhD; Bruce Hanik[2], PhD; Beth Chaney[1], PhD; Don Chaney[1], PhD; Bethany Tennant[1], MS; Enmanuel Antonio Chavarria[1], MS

[1]Department of Health Education and Behavior, University of Florida, Gainesville, FL, United States

[2]Department of Health and Kinesiology, Texas A&M University, College Station, TX, United States

Corresponding Author:
Michael Stellefson, PhD
Department of Health Education and Behavior
University of Florida
PO Box 118210
Gainesville, FL, 32611
United States
Phone: 1 352 294 1805
Fax: 1 352 392 1909
Email: mstellefson@ufl.edu

Abstract

Background: eHealth literacy refers to the ability of individuals to seek, find, understand, and appraise health information from electronic resources and apply such knowledge to addressing or solving a health problem. While the current generation of college students has access to a multitude of health information on the Internet, access alone does not ensure that students are skilled at conducting Internet searches for health information. Ensuring that college students have the knowledge and skills necessary to conduct advanced eHealth searches is an important responsibility particularly for the medical education community. It is unclear if college students, especially those in the medical and health professions, need customized eHealth literacy training for finding, interpreting, and evaluating health- and medical-related information available on the Internet.

Objective: The objective of our review was to summarize and critically evaluate the evidence from existing research on eHealth literacy levels among college students between the ages of 17 and 26 years attending various 4-year colleges and universities located around the world.

Methods: We conducted a systematic literature review on numerous scholarly databases using various combinations of relevant search terms and Boolean operators. The records were screened and assessed for inclusion in the review based on preestablished criteria. Findings from each study that met inclusion criteria were synthesized and summarized into emergent themes.

Results: In the final review we analyzed 6 peer-reviewed articles and 1 doctoral dissertation that satisfied the inclusion criteria. The number of participants in each reviewed study varied widely (from 34 to 5030). The representativeness of the results from smaller studies is questionable. All studies measured knowledge and/or behaviors related to college student ability to locate, use, and evaluate eHealth information. These studies indicated that many college students lack eHealth literacy skills, suggesting that there is significant room for improvement in college students' ability to obtain and evaluate eHealth information.

Conclusion: Although college students are highly connected to, and feel comfortable with, using the Internet to find health information, their eHealth literacy skills are generally sub par. College students, especially in the health and medical professions, would be well served to receive more customized college-level instruction that improves general eHealth literacy.

(J Med Internet Res 2011;13(4):e102) doi:10.2196/jmir.1703

KEYWORDS

eHealth literacy; college students; health occupations; professional preparation

XSL•FO
RenderX

Introduction

Electronic resources increasingly play a major role in consumer health, with the Internet being the preferred primary telecommunications vehicle for seekers of novel and germane health information. Although now widely relevant, the term electronic health information, also called *eHealth*, first appeared in 2000 to describe where health informatics, public health, health services, and information transmission processes intersected, primarily through Web-based applications [1,2]. Health information is one of the most investigated topics online [3]: 8 out of 10 Internet users report that they have at least once looked online for health information, making it the third most popular Web activity next to checking email and using search engines in terms of activities that almost everybody has done [4]. The importance of the Internet to acquire health information has spurred the creation of numerous eHealth information resources that assist consumers in discovering knowledge that can help promote and sustain personal health. Subsequent studies examining the effectiveness of eHealth interventions have proposed many definitions for eHealth [1,2,5]. Broadly stated, eHealth can also be thought of as the field where information and communication technology design enables the delivery of health-related and medical information [6]. While eHealth can potentially revolutionize medical and public health practice [7], numerous human resource, organizational, and cultural changes are still necessary to enable mainstream adoption of eHealth strategies for retrieving good-quality health information [1,8,9].

eHealth and the topic of *health literacy* are closely connected in public health. *Health literacy* is defined as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" [10]. It has been identified as a public health goal for the 21st century and stands as a significant challenge facing health care globally [11,12]. According to Norman and Skinner, the articulation of health literacy "underscores the importance of contextual factors that mediate health information and the need to consider health literacy in relation to the medium by which health resources are presented" [13]. The pervasiveness of the Internet has made obtaining, processing, and understanding health information using Web-based technologies a critical competency area for medical professionals. With the emergence of electronic medical and health records, medical mobile apps, and other related health informatics technologies, medical professionals are increasingly responsible for finding and evaluating health information resources electronically. In light of this, *eHealth literacy* now exists as an important skill set for health professionals tasked with seeking valid and reliable health information in a Web-based environment. However, most studies on literacy and health, such as the US Institute of Medicine's report titled *Health Literacy: A Prescription to End Confusion*, exclusively examine the relationship between health outcomes and literacy in the context of paper-based resources, not literacy in electronic environments [11,13]. Therefore, *eHealth literacy* is still a novel concept with varied definitions and models.

eHealth literacy refers to the ability of individuals to seek, find, understand, and appraise health information from *electronic resources* and apply such knowledge to addressing or solving a health problem [13]. eHealth literacy combines six core skills or types of literacy: traditional literacy, health literacy, information literacy, scientific literacy, media literacy, and computer literacy [13,14]. Table 1 [12,13,15,16] provides definitions of each type of literacy considered within the scope of eHealth literacy. These six facets have been developed by Norman and Skinner and have been depicted as the *eHealth Literacy Lily Model*, characterizing the six types of literacy as forming overlapping lily petals that feed into the overall eHealth literacy "pistil" (ie, center of the model). More specifically, the lily model categorizes the six core literacies into two primary types: analytic (ie, traditional, media, and information) and context- specific (ie, health, scientific, and computer). Analytic literacies refer to a set of skills that can be applied to an array of information sources, whereas context-specific literacies involve skills that are specific to a certain problem or situation. eHealth literacy, as the composite of both analytic and context-specific skills, requires the behavioral capability to do the following: work with technology, critically think about issues of media and science, and navigate through the vast array of eHealth decision-making resources. A variety of competencies are associated with obtaining eHealth information, including the knowledge, skills, abilities, and other attributes necessary to (1) conduct basic and advanced information searches, (2) apply Boolean operators to limit searches, (3) differentiate between scholarly documents, authoritative sources, periodicals, and primary sources of information, and (4) understand sometimes ambiguous eHealth terminology. Specific techniques using these proficiencies are necessary to find documents on the Web such as abstracts, bibliographies, research articles, and government reports. To ensure that individuals are optimally making use of available eHealth access, it is important that appropriate search-related practices and procedures be used to retrieve and assess the eHealth information that is located.

XSL•FO
RenderX

Table 1. Six components of eHealth literacy[13]

Type of literacy	Definition
Traditional literacy	Involves basic literacy skills, such as reading text, understanding written passages, and coherently speaking and writing a language [15].
Information literacy	According to the American Library Association, involves a person knowing "how knowledge is organized, how to find information, and how to use information in such a way that others can learn from them" [16].
Media literacy	Involves the ability to critically think about media content, and "enables people to place information in a social and political context and to consider issues such as the marketplace, audience relations, and how media forms in themselves shape the message that gets conveyed" [13].
Health literacy	Defined by the American Medical Association as a person's capability to "perform basic reading and numerical tasks required to function in the health care environment. Patients with adequate health literacy can read, understand, and act on health care information" [12].
Computer literacy	Involves the ability to use computers to solve problems. According to Norman and Skinner, "computer literacy includes the ability to adapt to new technologies and software and includes both absolute and relative access to eHealth resources" [13].
Scientific literacy	Involves an "understanding of the nature, aims, methods, applications, limitations, and politics of creating knowledge in a systematic manner" [13]. Allows health research findings to be placed in the appropriate context and requires the understanding of the discovery process.

Access to eHealth information is ubiquitous now for many who have broadband Internet; however, access to eHealth resources does not inevitably assure acuity in discerning good-quality health information from quackery on the Internet. The ability to diagnose and engage useful eHealth information from reputable medical sources, such as governmental agencies (eg, National Institutes of Health, Centers for Disease Control and Prevention, Health Canada) and medical establishments (eg, Mayo Clinic, WebMD, Canadian Medical Association) as compared with opinion or advertisements from so-called experts such as private sector marketers and nonverified public commentators, is becoming increasingly important. With the wealth of health information that exists on the Internet, this complex task requires far more interpretive and demonstrative skill than simply being able to enter a medical condition or term into an Internet search engine such as Google or Bing. For example, when using the Internet as a medical education resource, consumers should know how to critically examine and discriminate between primary and secondary sources of health information posted on a website [13].

Implementing effective Internet searches to locate health information is especially important for college students, as the Internet is now a favorite resource for information gathering among the "Millennial" generation. For the Millennial generation of college students, the Internet is a preferred source of health information [17]. While it may be safe to assume that college students have ample access to Web-based portals leading to eHealth information, it is important to be cognizant that access alone does not ensure that college students are adroit at searching for, locating, and evaluating health information. Ensuring that college students have the knowledge and skills necessary to conduct advanced eHealth searches is an important responsibility particularly for the medical education community.

To determine eHealth literacy among college students, it is first important to define the specific knowledge, skills, abilities, and other user attributes that have been considered in previous eHealth literacy research. Some of these attributes have been investigated by Ivanitskaya and Casey, who used the Information Literacy Competency Standards for Higher Education, developed by the Association of College and Research Libraries [18,19], to create the Research Readiness Self-Assessment (RRSA). The RRSA measures basic information literacy skills related to research ability. Specific to measuring information literacy, the competency model assesses knowledge and skill sets necessary to locate good-quality information on a specific health topic. These competencies verify abilities to determine possible sources of good-quality health information, conduct health information searches, evaluate the quality of the information, and appropriately use the information. Declarative knowledge, such as knowledge of plagiarism, health-related information sources, and research terminology, consists of typical knowledge variables measured in this competency-based approach. In addition, procedural knowledge, which involves skills and problem solving, includes knowledge of the procedures used to complete an information-seeking task electronically (ie, database navigation). Both types of knowledge are important for assessing the behavior and eHealth literacy of health information consumers [19].

In summary, there is a growing interest in eHealth literacy as an essential skill for students, especially those in the medical and health professions, and it is unclear whether the current level of eHealth literacy is sufficient, or whether customized eHealth literacy training for finding, interpreting, and evaluating health- and medical-related information available on the Internet at the college level would be required to nurture these skills. The purpose of this systematic review is to evaluate the current literature to determine whether college students can generally be considered an "eHealth literate" population.

Methods

Search Procedures

This review adopted the widely accepted definition of *eHealth literacy* as the ability of individuals to seek, find, understand, and appraise health information from electronic sources and apply such information to addressing or solving a health problem

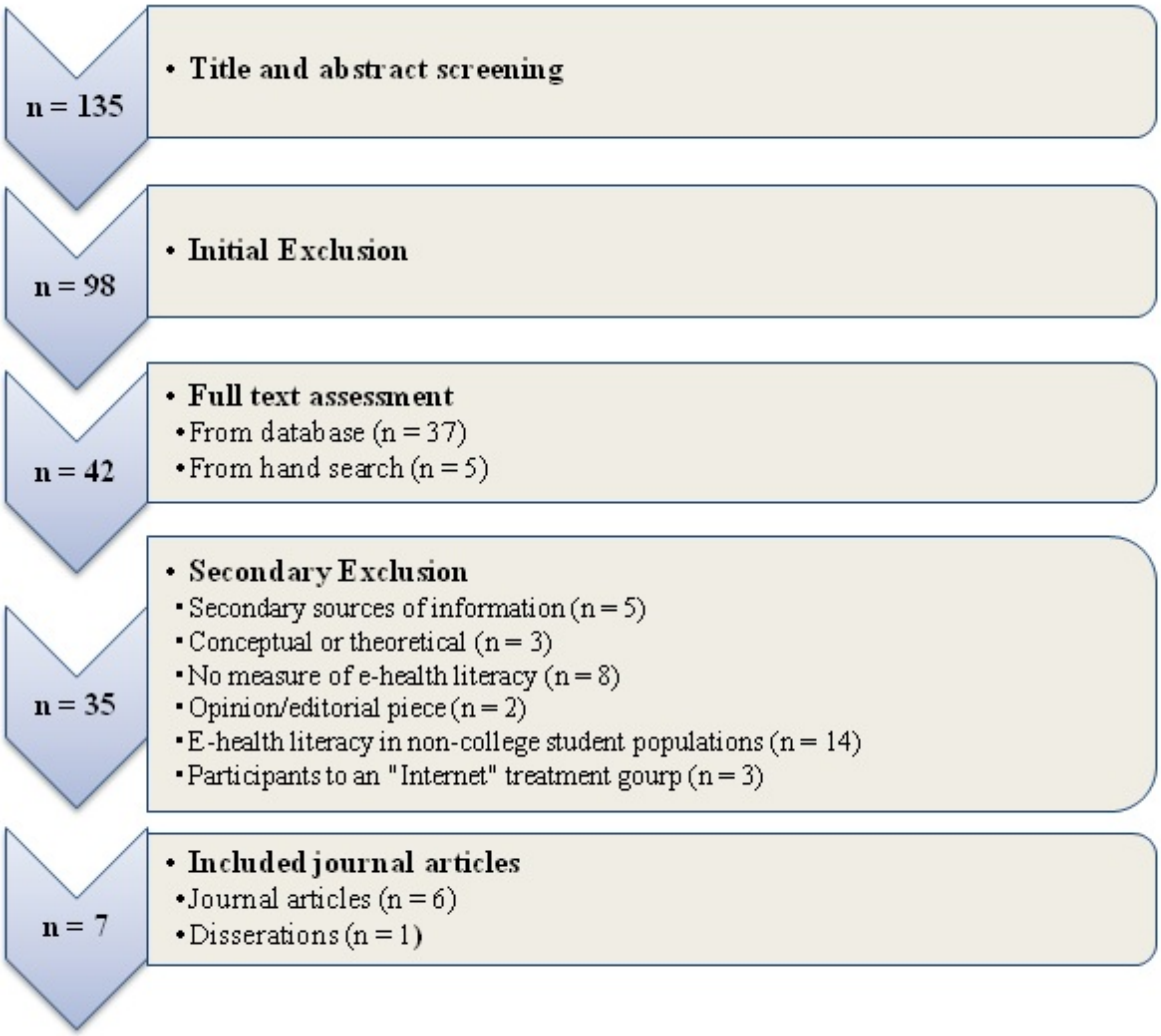
XSL•FO
RenderX

[13]. For this review, the experimental units of analysis for inclusion were peer-reviewed articles evaluating eHealth literacy (ie, seeking, finding, understanding, and appraising health information among electronic sources, primarily the Internet) exclusively among college students. The scope of the review was male and female college students between the ages of 17 and 26 years attending various 4-year colleges and universities located around the world. To generate a sample of empirical studies, we conducted an exhaustive search of electronic databases. Due to the relatively recent emergence of eHealth in the 21st century, only articles published from 2000 to the present day were eligible for inclusion. The actual search of all relevant literature took place during the spring of 2011. The searched databases were ERIC, PsycINFO, HealthSource, Medline, MasterFILE Premier, Academic Search Complete, CINAHL Plus with Full Text, Health Source: Nursing/Academic Edition, Psychology & Behavioral Sciences Collection, Applied Social Sciences Index and Abstracts, and CSA. The following key terms were entered in various combinations with multiple Boolean operators: *eHealth*, *electronic health*, *eHealth literacy*, *electronic health literacy*, *health literacy*, *Internet literacy*, *Internet health*, *electronic literacy*, *college students*, *university students*, and *literature review*.

All articles gathered through this initial search and screen process (n = 135) were evaluated for inclusion in the sample pool. We excluded 98 records after the screen of titles and abstracts. We initially excluded studies that did not survey 4-year college student populations between the ages of 17 and 26 years, and eliminated those that did not measure knowledge, skills, abilities or other attributes associated with eHealth literacy. In addition to the 37 papers that remained after the initial exclusion, we identified 5 other articles by hand searches after scanning the reference section of each database-identified article to enhance the breadth of the examination. This hand search resulted in the addition of 5 other articles meeting criteria for a full-text assessment. Overall, 42 papers were included in this full-text assessment, of which 35 were excluded for a variety of reasons, including (1) being secondary sources of information (n = 5), or purely conceptual or theoretical in scope (n = 3), (2) acting as opinion or editorial pieces (n = 2), (3) including populations other than college students (n = 14), (4) not explicitly measuring and reporting students' ability to seek, find, or evaluate electronic sources of health information (n = 8), or (5) reporting studies that assigned participants to an "Internet" treatment group within an intervention or trial (n = 3). After accounting for conditions outlined by the above exclusion criteria, we were left with 28 articles out of the review, leaving 7 articles that were empirical studies assessing eHealth literacy among college students. Figure 1 presents a flow diagram of the systematic literature review search process described above. Of the final 7 articles, 6 studies were carried out in the United States and 1 in Finland.

XSL•FO
RenderX

Figure 1. Flow diagram of the article selection process.



Methodological Data Analysis

To evaluate the methodological quality of each retained study, we used a modified version of criteria established by Nagel Bernstein and Freeman [20] to develop a methodological data score (MDS) for each article ranging from 0 (low) to 4 (high). If a study used multivariate procedures such as discriminant analysis, factor analysis, cluster analysis, hierarchical regression, or multivariate analysis of variance, then it received a score of 4. Articles reporting descriptive statistics, univariate regressions, or nonparametric tests such as chi-square were assigned a 3. Those reporting strictly qualitative data received a score of 2, and purely narrative descriptions or written observations received a score of 1. When studies did not report any statistical analysis procedures, then no points were awarded.

Results

Study Characteristics

Although eHealth has been a topic of interest since the turn of the century, this systematic search identified only 6 peer-reviewed articles [19,21-25] and 1 doctoral dissertation [25], published between 2005 and 2010. Four different journals published the articles: *Journal of Medical Internet Research* (2 articles), *Journal of American College Health* (2), *BMC Medical Informatics and Decision Making*, and the *Californian Journal of Health Promotion*. A total of 4 articles [19,21-23] explicitly defined eHealth literacy, but all explored at least one aspect of eHealth literacy accounted for in the Norman and Skinner [13] definition used within this review. For example, Nsuangani and Perez [21] asked specific questions about Internet use tendencies to find health information, while the RRSA, administered in 3 studies [19,25,26], sought to evaluate all dimensions of eHealth literacy.

The studies included in this systematic review were generally exploratory in nature. Many used demographic variables to group students in a nonexperimental fashion and then explored differences in relationships. The independent variables used in all studies were unique and directly related to the study purpose; however, the most common independent variables were users and nonusers of the Internet, gender, student classification, and race. The dependent variables included self-reported use of Web-based health advice services [22]; perceptions of the accuracy of health information on the Internet [21]; perceptions of the privacy of health information on the Internet [21]; frequency of Internet use for seeking health information [20,22,23]; quality of health-related websites [19,25,26];

XSL•FO
RenderX

attitudes and beliefs about using the Internet for finding health information (eg, beliefs that open-access Internet and search engines are always the best sources of information) [25]; ability to find electronic health information [19,25,26]; ability to evaluate electronic health information [19,25,26]; perceived research skills (ie, self-reported subjective beliefs about one's own skills) [19,25,26]; ability to judge the trustworthiness of Internet pharmacies [25]; and number of correct answers to sexual health questions following the conduct of Internet searches [24].

Results from the methodological assessment described above indicated that the reviewed articles had similar degrees of analytic rigor. That is, 2 studies randomly assigned participants to treatment groups [21,22], while the others used convenience [19,23,25,26] and purposive [24] sampling techniques; 6 of the studies [19,21-23,25,26] used a quantitative paradigm to determine patterns between independent and dependent variables, and 1 study [24] used a mixed-methods approach. Only 2 studies [21,26] used validated surveys containing reliability estimates for the data collected, while 3 studies [22,25,26] did not explicitly report this information. A total of 3 studies [21-23] used chi-square as the analysis of choice to explore differences in patterns between groups, and the remaining studies [19,24-26] simply reported descriptive statistics. Redmond [26] used multiple *t* tests to determine differences in eHealth literacy skills between rural and nonrural college students. Ivanitskaya and Brookins-Fisher [25] performed multiple independent *t* tests to assess whether differences in critical judgment existed among students who either did or did not use the Internet for health decision making. The mean MDS for the reviewed studies was 3.14 (SD 0.38), with 6 of the 7 studies (86%) scoring 3. Table 2 describes the basic design, measurement, and analysis of each study accompanied by each study's individual MDS.

Table 2. Design, measurement, analysis, and methodological data score (MDS) of selected studies

First author (year)	Sample size	Design	Instrument	Instrument validity	Instrument reliability	Analysis	MDS
Ivanitskaya [19] (2006)	308	Nonexperimental	RRSA[a]	Face, content	Yes, but no value reported	Descriptive statistics, multiple regression	3
Nsuangani [21] (2006)	136	Experimental	Ad hoc survey	Face (expert panel)	Pre–post, κ = 0.41 for items retained for analysis	Frequency distributions; cross-tabulations; chi-square	3
Castren [22] (2008)	5030	Experimental	Student Health Survey 2004 (Finland)	Not reported	Not reported	Frequency distributions; cross-tabulations; chi-square	3
Escoffery [23] (2005)	743	Nonexperimental	Ad hoc survey	Not reported	Not reported	Descriptive statistics; chi-square	3
Buhi [24] (2009)	34	Nonexperimental	Ad hoc	Content (implied)	Not reported	Descriptive statistics	3
Ivanitskaya [25] (2010)	1914	Nonexperimental	RRSA	Face, content	Not reported	Descriptive statistics; probabilities; *t* tests; hierarchical regression analysis	4
Redmond [26] (2007)	243	Nonexperimental	RRSA	Face, content	Ability to obtain health information, alpha = .69; ability to evaluate electronic health information, alpha = .65; overall health information competency, alpha = .77	Descriptive statistics; *t* tests; Cohen d	3

[a] Research Readiness Self-Assessment.

Demographics

The number of participants in each reviewed study varied widely (ranging from 34 to 5030), which calls into question the representativeness of the results from those studies with smaller samples. Additionally, the research findings related to gender varied within 5 studies [19,21-23,26]. For example, Nsuangani and Perez [21] found that male college students were more likely to use the Internet to buy pharmaceutical products and locate consumer health information, whereas female students were more likely to obtain general health- and medical-related information online. Interestingly, this finding was supported in 2 other studies as well [22,23]. Males were more likely to seek out medical consultations using the Internet [21], while females were more likely to self-report diagnosing chronic health conditions using the Internet [22]. Based on this evidence, it appears that females used the Internet more for health information and diagnostic purposes, while males were more likely to use Internet for consumer health products and services. Also, male and female college students did not differ significantly as to whether they expressed concern regarding the accuracy of health information found on the Internet [21]. We found no statistically significant differences on any eHealth literacy outcome when considering race or ethnicity.

Obtaining Health Information Using the Internet

We found that in 3 studies performed in the United States exploring the percentage of college students using the Internet to acquire health information, 91 of 136 (67%), 549 of 743 (73.9%), and 24 of 34 (71%) college students surveyed had ever

XSL•FO
RenderX

used the Internet to search for health information. In 1 study, 111 of 743 (14.9%) college students reported using the Internet to locate health information in the past day or week, with less than one-third reporting doing so in the past month [23]. In the same study, 539 of 743 (72.5%) students reported being averse to logging onto a health program delivered over the Internet, and only 204 of 743 (27.5%) students surveyed reported willingness to participate in a health program on the Internet. Another study reported that participants were reluctant to use the Internet for interactive health purposes, with 119 of 136 surveyed students (87.5%) reporting an unwillingness to use online medical discussion applications [21]. Another study conducted in Finland corroborated this reticence to participate in online health programming, finding that only 370 of 3153 (11.7%) Finnish undergraduate students had ever used a Web-based health advice service offered to them through their student health services department [22].

While 1 study [23] suggested that 393 of 743 (52.9%) college students surveyed would like to individually retrieve health information on the Internet, several studies indicated college students self-reported a lack of skills necessary to execute successful health-related searches on the Internet [19,23,25,26]. Escoffery et al [23] noted that 661 of 743 (89.0%) college students surveyed did not always find their desired eHealth information. Among those, 82 (11%) students surveyed did feel that they were capable of finding health information on the Internet, whereas only slightly more than half reported success "most of the time." Two studies [19,26] determined that many college students are rather unsophisticated health information seekers when using the Internet. Another study noted that college students were unable to critically evaluate health information found on the Internet [25]. Students were also unaware of the difference between a primary and secondary source of data when attempting to locate online journal articles in the health-related fields [19,26]. Finally, students who used eHealth information to help make health decisions had lower overall critical judgment ability than those who used nonelectronic sources of information for the same purpose [25].

Perceived versus Actual eHealth Literacy

Ivanitskaya et al [19] and Redmond [26] assessed (1) how students felt about their own level of eHealth literacy, (2) how proficient students were at searching for and evaluating eHealth information, and (3) how well students understood the difference between peer-reviewed scholarly resources and opinion pieces or sales pitches. Both studies used the RRSA online assessment tool, which evaluated perceived and actual knowledge of student ability in browsing the Internet and researching health topics given various search scenarios. The RRSA, based on the Information Literacy Competency Standards for Higher Education, assesses knowledge and skills related to locating, evaluating, and using good-quality sources of eHealth information. Specifically, the RRSA contains the following items: "(1) multiple choice or true/false questions that measure declarative knowledge; (2) interactive, problem-based exercises that measure procedural knowledge; (3) demographic questions; and (4) a question that asks for a self-report about the level of the respondent's research skills" [19]. For example, a knowledge-based item in the survey asks respondents to indicate which Boolean operator (eg, "and," "or," or "not") produces the most Internet search results (answer: or). An example of a skill-based survey item asks respondents to determine which Boolean operator is appropriate for a particular search situation, then requests that the respondent perform an Internet search using that particular Boolean operator, followed by reporting back the number of Web hits generated by the search [27]. There are two subscale measurements within the RRSA: Actual Ability to Obtain (AAO) eHealth information and Actual Ability to Evaluate (AAE) eHealth information. The AAO subscale comprises 11 multiple choice items where total scores can range from 0 to 16. The AAE subscale comprises 13 multiple choice items where total scores can range from 0 to 23. A higher score on both subscales indicates better actual ability. One study within this review [19] demonstrated that the data derived from the RRSA possessed satisfactory internal reliability (alpha = .78). Ivanitskaya et al [19] found that 258 of 306 (84%) college students surveyed perceived their eHealth literacy skills as "good," "very good," or "excellent," yet students' scores on a 56-item scale evaluating their actual eHealth literacy skills were very poor (mean 37%, SD 6.4%).

Also, it was found that within each perceived skill category (eg, perceived ability to find health information and perceived ability to judge the quality of health information), the actual overall competency scores of college students varied greatly. Specifically, the ability of college students to evaluate their own competency was inconsistent with their actual eHealth literacy. Redmond [26] found that nonrural college students were better able to obtain eHealth information than were rural college students, but there were no statistically significant differences in the ability to evaluate eHealth information between the two groups. Escoffery et al [23] found that 260 of 743 (35%) college students surveyed expressed "serious concern" about their ability to find good-quality health information using the Internet, while only a small proportion, 52 of 743 (7%), expressed "no concern" regarding the accuracy of health information they acquired on the Internet. Despite the relatively higher level of apprehension regarding ability to find eHealth information, 204 of 514 (39.7%) college students who reported seeking health information online believed that being able to retrieve health information online improved the way they took care of their health "some" or "a lot."

In light of these findings, all studies tended to agree that college students in general [19,21,23-26], and those in health and medical professional programs specifically [19,26], should further develop their proficiency in appraising, using, and evaluating health information found on the Internet. Table 3 describes the primary findings gathered from the research questions posed in each study.

XSL•FO
RenderX

Table 3. Primary findings from research questions

First author	Research question(s)	Findings
Ivanitskaya [19]	How proficient are university students at searching for health-related information?	Students are not proficient at advanced health information searches.
	How proficient are university students at evaluating health-related information?	Students have mixed proficiency at evaluating health-related information.
	How well do university students understand the difference between peer-reviewed scholarly resources, opinion pieces, and sales pitches?	Students are deficient in discriminating between different types information sources.
	How aware are university students of their own level of health information competencies?	Undergraduate students are inaccurate judges of their own health information competencies. Self-reports may not accurately predict students' actual health information competencies.
Nsuangani [21]	Do male and female college students differ in their Internet behaviors related to health?	Males more likely than females to report online medical consultation. Males are more likely to buy pharmaceuticals online. More males use email to communicate with a health care provider.
Castren [22]	Does self-reporting of chronic conditions differ between users and nonusers of a Web-based health advice service?	Male users of a health advice service had a higher rate of self-reported chronic conditions than male nonusers; female users of a health advice service had a higher rate of a reported chronic condition than female nonusers.
Escoffery [23]	Are there differences in Internet use for health information by level of Internet experience?	There is no difference.
	Are there differences in Internet use for health information by gender?	Significantly more female than male students obtain health information online.
	Are there differences in Internet use for health information by level in college?	There is no difference.
Buhi [24]	When asked questions about sexual health, do college students find accurate answers online?	For 12 of the 13 questions asked, at least 24 of 34 (71%) students answered the questions correctly. Of 34 students surveyed, 17 (50%) correctly answered the question that asked them to locate an anonymous HIV test in the local area.
Ivanitskaya [25]	To what degree are college-educated information seekers able to determine trustworthiness of online pharmacies?	How college students rate trustworthiness of online pharmacies varies substantially. Only 593 of 1914 (31.0%) respondents gave low ratings to untrustworthy online pharmacies.
	Do those who used information to make health decisions have better judgment skills?	Respondents using online health information for decision-making have significantly worse judgment than those not using online health information for decision-making.
Redmond [26]	Do rural and nonrural freshmen differ in their ability to obtain health information?	A statistically significant difference exists, with nonrural students performing higher than rural students, $t_{241} = 2.23$, $P = .03$, Cohen d = .29.
	Do rural and nonrural freshmen differ in overall health information competency?	No difference exists, $t_{241} = –.14$, $P = .89$, Cohen d = .02.
	Do rural and nonrural freshmen differ in their ability to evaluate health information?	No difference exists, $t_{241} = 1.34$, $P = .18$, Cohen d = .18.

Discussion

Main Findings

The main conclusion of this systematic review was that college students may lack important skills for seeking and evaluating health information available on the Internet. While college students, for the most part, have convenient access to health information on the Internet, this systematic review indicated that many students possess weak eHealth literacy skills related to searching for, retrieving, using, and evaluating sources of eHealth information. Furthermore, 3 studies [19,25,26] noted that the subjective self-perceptions of college students regarding their ability to use eHealth information sources were incongruent with their demonstrated eHealth literacy skills. Therefore, it is possible that college students may be mistakenly judging their own ability to successfully locate and evaluate eHealth information. They may (or may not) hold an overly optimistic view of their ability to do Internet research on health-related topics. While it is clearly too early in this field of investigation to state definitively that there is a gap between perceived and actual eHealth literacy among college students, the trend noted in this systematic review provides impetus for future research to either support or disconfirm whether this phenomenon may truly exist.

Regardless of whether a discord exists between perceived and actual eHealth literacy among college students, there nevertheless is an invaluable opportunity to build medical education competencies among college-age students, especially

XSL•FO
RenderX

those seeking degrees in the medical and health professions. College students surveyed in the reviewed studies did not achieve satisfactory levels of eHealth literacy; thus, we should perhaps reexamine the standards that are being used to measure eHealth literacy among this diverse audience. While valid and reliable *health* literacy measures have been widely established, far fewer instruments are universally accepted as accurately and appropriately assessing *eHealth* literacy. It is possible that the current standard being promulgated might be appropriate only for technologically elite audiences. Supposing that consumers will meet eHealth literacy standards set by technicians is probably unrealistic. High-stakes measures used in the current studies may have attempted to assess skill navigating the Internet to locate health information, but these measures may do so in a manner less applicable to a wide-ranging audience of future public health professionals. Future professionals, especially in a health-related field. will undoubtedly be using the Internet and related health informatics technologies to gather, manage, and deliver health information; however, we have yet to fully understand the context of the interactions occurring between diverse users and health informatics technologies (given that the consumer health informatics field is still in its infancy). There is a strong possibility that the broad-based, multidimensional definition of eHealth literacy is overly ambitious, even for individuals seeking an advanced degree in a health-related field.

While measurement issues are important to consider, it is also important to recognize that future inquiries should avoid reporting purely descriptive self-report data on frequency of use and self-efficacy for using the Internet to find health information. Data from self-reports depicting college students as both frequent and confident eHealth users may be more assumptive than truly substantive, especially considering the current research. As explained by Bandura within self-efficacy theory, “expectation alone will not produce desired performance if the [individual’s] component capabilities are lacking” [28]. Thus, what might be more important is testing of relevant skills in this new area of inquiry. Thus, more research should evaluate the most effective instructional strategies for molding able-bodied “eHealth educators” within a variety of medical and allied health professional preparation programs. Planned instructional experiences must consider the unique eHealth literacy competencies that are expected of college students studying to become health professionals, a distinction that places them in a unique position as compared with the general public. These students are expected to gravitate toward evidence-based practice and to critically appraise qualified sources of health information using specific resources such as The Cochrane Collaboration [29] or the Guide to Community Preventive Services [30]. More studies of college students at varied institutions, majoring in a variety of health and medical programs, will enable the eHealth literacy research community to ask and answer more targeted research questions with more specific audiences.

Demographics

The literature also indicated a tendency for male college students to be more likely to use the Internet to locate and acquire consumer health products (eg, pharmaceuticals, dietary/sports supplements, vitamins, performance-enhancing substances) and services (eg, Web-based medical consultations) and less likely to search for general information on illness, disease, or disease prevention using medical reference websites. Female college students, on the other hand, were generally more likely to conduct these types of general health or medical searches and were less likely to obtain health services over the Internet (eg, accessing primary care physicians’ Web portals, communicating by email with health care providers). In light of this interesting preliminary trend, future research would benefit from further study regarding what particular Internet search and retrieval characteristics can be attributed to male or female college students. Unique search propensities could speak to various developmental issues of marketing pressures, peer influences, and even health privacy concerns.

Obtaining Health Information Using the Internet

While the literature supports college students wanting to use the Internet to seek out general health information, there is little evidence to suggest that students care to discuss their own health problems or obtain personalized medical advice over the Internet. College students reported reluctance to using interactive Internet applications for health communication purposes (ie, electronic communication with health care providers). This finding revealed itself not only in the United States, but also in 1 Finnish study that we reviewed. Among college students, the convenience of using the Internet for *seeking* personal health information may be valued more so than the prospect of *receiving* individualized feedback on personal health concerns or problems via interaction with a qualified medical professional. This could be the result of contextual Web security issues affecting confidentiality. The issue of trust when using the Internet to seek and share medical information is an important one to consider, especially with the emergence of peer-to-peer or horizontal health communication among college students. More research should be done to discover what particular sources of Web-based health information college students are consulting and which cause uneasy feelings originating from potential threats to data security and privacy.

Limitations

This systematic review had several limitations. Although we conducted a comprehensive literature search on numerous databases using a variety of pertinent search terms, certain studies may have been overlooked due to lack of indexing in searched databases. In addition, all studies were carried out in either the United States or Finland, which are both highly technologically savvy countries. Also, Finland is regarded as one of the world’s most literate societies, with high levels of educational attainment [31]. Another noted limitation is that one standard definition of eHealth does not exist, which limits the ability of researchers to find all articles examining eHealth literacy within a single literature review. Another limitation involves the number of articles included in the review. Although the studies reached similar conclusions in selected instances, the small sample of studies (n = 7) may not truly reflect the population’s (ie, college students) true eHealth literacy levels. In addition, most studies used convenience sampling techniques, which can result in findings not being reflective of the true

XSL•FO
RenderX

populations of interest. As well, most studies in this review (n = 4) collected self-report data and did not test actual eHealth literacy skills to complement students' self-perceptions.

The marketplace penetration of information technology into college students' lives and educational settings is shifting rapidly (eg, smartphones, social networking websites, iPads). The reviewed studies of eHealth literacy among college students did not distinguish these emerging applications among the many alternative electronic sources of information, which may not truly reflect current search tendencies of college students. These types of applications conducive to mobile health information searches have spawned the new field of mHealth, which may suggest broadening or revising the study of eHealth literacy among college students. Finally, while the mean MDSs for the studies in this systematic review were quite good, few reported sufficient validity and reliability measures for data collected with survey or testing instruments, and almost all data analyses were univariate versus multivariate.

Comparison With Prior Works

Even where access to basic Internet infrastructure exists or is provided, optimal Internet use is often limited by other factors, such as human interface. To some extent, human interface encompasses issues commonly considered when assessing *usability*. Usability of an eHealth information source typically refers to the quality of a user experience when interacting with the resource, with an emphasis on behavior rather than opinion or recollection [32,33]. The construct measures learnability, memorability, efficiency, frequency, and severity of errors. All of these aspects are affected by human limitations, such as literacy, and by health website quality criteria, such as accuracy, completeness, readability, and design. Thus, the construct of usability is inextricably linked with eHealth literacy. There are varying levels of usability among eHealth resources, so it would be useful to determine whether the perceived usability of resources is related to eHealth literacy outcomes [34,35]. An analysis that assesses individual perceptions of eHealth usability in relation to overall behavioral capability to locate and evaluate eHealth information is vital for future eHealth literacy research [33,35]. Studying consumer health informatics (ie, analyzing consumer needs for acquiring and using information retrieved using technology) in conjunction with eHealth literacy [34,35] can further develop methods that pave the way for providing health care service in the information age.

Consequently, collegiate degree programs for those entering the medical and allied health fields are uniquely positioned to nurture and develop eHealth competencies among future health professionals. It is important for education administrators to determine (1) what list of eHealth topics should be covered, (2) what types of courses and materials can address the needed competencies, (3) how many hours of subject matter instruction might be necessary for eHealth literacy skill development, and (4) whether eHealth warrants a specific emphasis area or track within professional preparation programs. Creating mission and policy statements that give attention to these relevant aspects of eHealth literacy instruction will help improve student outcomes.

Conclusions

Evidence from this systematic review suggests that future health professionals need professional preparatory experiences that help build their eHealth literacy proficiencies. Enhanced skills development will likely develop as a product of practical medical Internet research opportunities that encourage critical thinking among students. As suggested by Escoffery et al [23], and supported by this systematic review, more needs to be done to inform the training of students in the health and medical professions, to "search the Internet for health information and to evaluate health information on Web sites." Because of this, two important research questions should continue to be investigated in medical education. First, do professionally prepared college students in the health professions have the skills to navigate electronic resources to retrieve evidence-based health information? Second, do college students studying to be health professionals have an inflated sense of self-efficacy regarding their actual ability to locate and evaluate good-quality health information on the Internet?

Given that governmental and advisory agencies have designated eHealth literacy as paramount to improving societal health in both Canada [36] and the United States [37], it is important that future eHealth educators be provided with planned learning experiences in this growing field. Several health and medical disciplines have recognized this need area and have incorporated formal professional responsibilities related to eHealth literacy into core competency development models. For example, future professionals in the field of health education are expected to find valid health information resources electronically and evaluate the usefulness of such information [38]. It is important that health and medical education programs develop these types of proficiencies among future health professionals. Both current and future college students, especially those in the medical and health professions, need customized eHealth literacy training for finding, interpreting, and evaluating health- and medical-related information available on the Internet.

Conflicts of Interest

None declared

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] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
2. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]

XSL•FO
RenderX

3. Taylor H, Leitman R. Healthcare News;1(13):1-3. 2001 Apr 23. eHealth traffic critically dependent on search engines and portals URL: http://www.harrisinteractive.com/news/allnewsbydate.asp?NewsID=270 [accessed 2011-05-27] [WebCite Cache ID 5z0FCTVgX]
4. Fox S. Health topics: 80% of Internet users look for health information online. Washington, DC: Pew Internet & American Life Project; 2011 Feb 01. URL: http://pewinternet.org/~/media//Files/Reports/2011/PIP_HealthTopics.pdf [accessed 2011-10-05] [WebCite Cache ID 62D0se6Gu]
5. Jones S, Johnson-Yale C, Millermaier S, Perez FS. First Monday(10). 2009. Everyday life, online: U.S. college students' use of the Internet URL: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2649/2301 [accessed 2011-10-10] [WebCite Cache ID 62L2Sp9eS]
6. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
7. Kwankam SY. What e-Health can offer. Bull World Health Organ 2004 Oct;82(10):800-802. [Medline: 15643805]
8. Kreps L, Neuhauser L. Ehealth and the delivery of health care. J Comput Mediat Commun 2010;15(3):364-366. [doi: 10.1111/j.1083-6101.2010.01524.x]
9. Kelley H, Chlasson M, Downey A, Pacaud D. The clinical impact of eHealth on the self-management of diabetes: a double adoption perspective. J Assoc Inf Syst 2011;12(3):208-234.
10. US Department of Health and Human Services. US Government Printing Office. Healthy People 2010: Understanding and Improving Health. 2nd edition URL: http://www.healthypeople.gov/2010/redirect.aspx?url=/2010/document/pdf/uih/2010uih.pdf [accessed 2011-07-22] [WebCite Cache ID 60NFpVRES]
11. Nielsen-Bohlman L, Panzer AM, Kindig DA. Health Literacy: A Prescription to End Confusion. Washington, DC: National Academies Press; 2004.
12. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs. Health literacy: report of the Council on Scientific Affairs. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. JAMA 1999 Feb 10;281(6):552-557. [Medline: 10022112]
13. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
14. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
15. Tyner K. Literacy in a Digital World: Teaching and Learning in the Age of Information. Mahwah, NJ: Erlbaum; 1998.
16. Presidential Committee on Information Literacy. American Library Association. 1989 Jan 10. Final Report URL: http://www.ala.org/ala/mgrps/divs/acrl/publications/whitepapers/presidential.cfm [accessed 2011-05-27] [WebCite Cache ID 5z0FaJ7f9]
17. Heuberger RA, Ivanitskaya L. Preferred sources of nutrition information: contrasts between younger and older adults. J Intergener Relatsh 2011;9(2):176-190. [doi: 10.1080/15350770.2011.567919]
18. Association of College & Research Libraries. American Library Association. 2011. ACRL Publications URL: http://www.ala.org/ala/mgrps/divs/acrl/publications/index.cfm [accessed 2011-05-27] [WebCite Cache ID 5z0GWX4Jv]
19. Ivanitskaya L, O'Boyle I, Casey AM. Health information literacy and competencies of information age students: results from the interactive online Research Readiness Self-Assessment (RRSA). J Med Internet Res 2006;8(2):e6 [FREE Full text] [doi: 10.2196/jmir.8.2.e6] [Medline: 16867969]
20. Nagel Bernstein I, Freeman HE. Academic and Entrepreneurial Research: The Consequences of Diversity in Federal Evaluation Studies. New York, NY: Russell Sage Foundation; 1975.
21. Nsuangani NM, Pérez MA. Accessing web-based health related information by college students: An exploratory study. Californian J Health Promot 2006;4(1):64-74 [FREE Full text]
22. Castrén J, Huttunen T, Kunttu K. Users and non-users of web-based health advice service among Finnish university students - chronic conditions and self-reported health status (a cross-sectional study). BMC Med Inform Decis Mak 2008;8:8 [FREE Full text] [doi: 10.1186/1472-6947-8-8] [Medline: 18237414]
23. Escoffery C, Miner KR, Adame DD, Butler S, McCormick L, Mendell E. Internet use for health information among college students. J Am Coll Health 2005;53(4):183-188. [doi: 10.3200/JACH.53.4.183-188] [Medline: 15663067]
24. Buhi ER, Daley EM, Fuhrmann HJ, Smith SA. An observational study of how young people search for online sexual health information. J Am Coll Health 2009;58(2):101-111. [doi: 10.1080/07448480903221236] [Medline: 19892646]
25. Ivanitskaya L, Brookins-Fisher J, O Boyle I, Vibbert D, Erofeev D, Fulton L. Dirt cheap and without prescription: how susceptible are young US consumers to purchasing drugs from rogue internet pharmacies? J Med Internet Res 2010;12(2):e11 [FREE Full text] [doi: 10.2196/jmir.1520] [Medline: 20439253]
26. Redmond TL. Electronic (Digital) Health Information Competency: A Comparative Analysis of Knowledge and Skills of Rural and Non-rural Freshman College Students dissertation. Mount Pleasant, MI: Central Michigan University; 2007.
27. Hanik B, Stellefson M. E-health literacy competencies among undergraduate health education students: A preliminary study. Int Electronic J Health Educ 2011;14:46-58 [FREE Full text]
28. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]

XSL•FO
RenderX

29. The Cochrane Collaboration. 2010. The Cochrane Library Database URL: http://www.cochrane.org/ [accessed 2011-07-09] [WebCite Cache ID 603vpBIuY]
30. Guide to Community Preventive Services. The Community Guide Branch, Epidemiology Analysis Program Office (EAPO), Office of Surveillance, Epidemiology, and Laboratory Services (OSELS), Centers for Disease Control and Prevention. 2011 Jun 30. The Community Guide URL: http://www.thecommunityguide.org/index.html [accessed 2011-07-09] [WebCite Cache ID 603wd3REz]
31. Darling-Hammond L, Rothman R. Alliance for Excellent Education and Stanford Center for Opportunity Policy in Education. editors. Teacher and Leader Effectiveness in High-Performing Education Systems URL: http://www.all4ed.org/files/TeacherLeaderEffectivenessReport.pdf [accessed 2011-07-22] [WebCite Cache ID 60NKrhsye]
32. Atkinson NL, Gold RS. The promise and challenge of eHealth interventions. Am J Health Behav 2002;26(6):494-503. [Medline: 12437024]
33. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
34. Stellefson M, Chaney B, Chaney D. The digital divide in health education: myth or reality? Am J Health Educ 2008;39(2):106-112.
35. Korp P. Health on the Internet: implications for health promotion. Health Educ Res 2006 Feb;21(1):78-86 [FREE Full text] [doi: 10.1093/her/cyh043] [Medline: 15994845]
36. Canadian Council on Learning. 2008. Health Literacy in Canada: A Healthy Understanding URL: http://www.ccl-cca.ca/pdfs/HealthLiteracy/HealthLiteracyReportFeb2008E.pdf [accessed 2011-07-22] [WebCite Cache ID 60NNZEBdF]
37. Anonymous. US Department of Health and Human Services, Office of Disease Prevention and Health Promotion. 2010. National Action Plan to Improve Health Literacy URL: http://www.health.gov/communication/hlactionplan/pdf/Health_Literacy_Action_Plan.pdf [accessed 2011-07-22] [WebCite Cache ID 60NO2GQ4U]
38. National Commission for Health Education Credentialing, American Association for Health Education, Society for Public Health Education. A Competency-Based Framework for Health Educators--2006. Whitehall, PA: National Commission for Health Education Credentialing; 2006.

Abbreviations

AAE: Actual Ability to Evaluate
AAO: Actual Ability to Obtain

Edited by G Eysenbach; submitted 02.12.10; peer-reviewed by L Ivanitskaya, E Neter, T Matingwina, M Stephens; comments to author 01.04.11; revised version received 02.06.11; accepted 19.07.11; published 01.12.11.

Please cite as:
Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA
eHealth Literacy Among College Students: A Systematic Review With Implications for eHealth Education
J Med Internet Res 2011;13(4):e102
URL: http://www.jmir.org/2011/4/e102/
doi:10.2196/jmir.1703
PMID:22155629

©Michael Stellefson, Bruce Hanik, Beth Chaney, Don Chaney, Bethany Tennant, Enmanuel Antonio Chavarria. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 01.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Analysis of 4999 Online Physician Ratings Indicates That Most Patients Give Physicians a Favorable Rating

Bassam Kadry[1], MD; Larry F Chu[1], MS, MD; Bayan Kadry[2], PharmD; Danya Gammas[3], BS; Alex Macario[1], MBA, MD

[1]Department of Anesthesia, School of Medicine, Stanford University, Stanford, CA, United States

[2]Eugene Applebaum College of Pharmacy, Wayne State University, Detroit, MI, United States

[3]School of Pharmacy, Creighton University, Omaha, NE, United States

Corresponding Author:
Bassam Kadry, MD
Department of Anesthesia
School of Medicine
Stanford University
300 Pasteur Drive
H3580
Stanford, CA, 94305
United States
Phone: 1 (650) 723 6415
Fax: 1 (650) 725 8544
Email: bkadry@stanford.edu

Abstract

Background: Many online physician-rating sites provide patients with information about physicians and allow patients to rate physicians. Understanding what information is available is important given that patients may use this information to choose a physician.

Objectives: The goals of this study were to (1) determine the most frequently visited physician-rating websites with user-generated content, (2) evaluate the available information on these websites, and (3) analyze 4999 individual online ratings of physicians.

Methods: On October 1, 2010, using Google Trends we identified the 10 most frequently visited online physician-rating sites with user-generated content. We then studied each site to evaluate the available information (eg, board certification, years in practice), the types of rating scales (eg, 1–5, 1–4, 1–100), and dimensions of care (eg, recommend to a friend, waiting room time) used to rate physicians. We analyzed data from 4999 selected physician ratings without identifiers to assess how physicians are rated online.

Results: The 10 most commonly visited websites with user-generated content were HealthGrades.com, Vitals.com, Yelp.com, YP.com, RevolutionHealth.com, RateMD.com, Angieslist.com, Checkbook.org, Kudzu.com, and ZocDoc.com. A total of 35 different dimensions of care were rated by patients in the websites, with a median of 4.5 (mean 4.9, SD 2.8, range 1–9) questions per site. Depending on the scale used for each physician-rating website, the average rating was 77 out of 100 for sites using a 100-point scale (SD 11, median 76, range 33–100), 3.84 out of 5 (77%) for sites using a 5-point scale (SD 0.98, median 4, range 1–5), and 3.1 out of 4 (78%) for sites using a 4-point scale (SD 0.72, median 3, range 1–4). The percentage of reviews rated ≥75 on a 100-point scale was 61.5% (246/400), ≥4 on a 5-point scale was 57.74% (2078/3599), and ≥3 on a 4-point scale was 74.0% (740/1000). The patient's single overall rating of the physician correlated with the other dimensions of care that were rated by patients for the same physician (Pearson correlation, $r = .73$, $P < .001$).

Conclusions: Most patients give physicians a favorable rating on online physician-rating sites. A single overall rating to evaluate physicians may be sufficient to assess a patient's opinion of the physician. The optimal content and rating method that is useful to patients when visiting online physician-rating sites deserves further study. Conducting a qualitative analysis to compare the quantitative ratings would help validate the rating instruments used to evaluate physicians.

(J Med Internet Res 2011;13(4):e95) doi:10.2196/jmir.1960

XSL•FO
RenderX

KEYWORDS

Doctor ratings; patient satisfaction; online physician reviews; consumer health; physician rating

Introduction

In 2010, 88% of adult Americans used the Internet to search for health-related information [1-3]. Patients are seeking information not only about disease conditions but also about physicians and hospitals. In fact, in the United States, 47% looked up information about their providers online, 37% consulted physician-rating sites, and 7% of people who sought information about their provider posted a review online [4]. A separate study found that 15% of consumers compare hospitals before making a selection, and 30% of consumers compare physicians online before making a selection [5].

Many physician-rating websites provide users with basic information about the physician such as years in practice and contact information [6,7]. Some of the websites access various databases to display further information about board certification, residency, and any disciplinary action [8]. This information can be obtained for free, or patients can pay to obtain a more in-depth report about the physician [9].

Many websites enable users to enter reviews and rankings about specific physicians. This capability has drawn the attention of consumer advocacy groups, providers, insurance companies, and hospitals. Although knowledge about the patient experience is useful, critics of these portals identify them as being at risk for misinformation, sabotage, and manipulation [10-14]. Few large-scale studies have been conducted to assess the content and rating methods of these physician-rating sites [15].

The goals of this study were to (1) determine the most frequently visited physician-rating websites that have user-generated content, (2) evaluate the content characteristics of each site to rate physicians, and (3) analyze online ratings of 4999 individual physician ratings.

Methods

Approval for this study was obtained from the Institutional Review Board at Stanford University School of Medicine.

The Most Commonly Visited Physician-Rating Sites

A search of the Internet (Bing, Google, Google Directory, Google Trends, Blekko, Yahoo, and Yahoo Directory) with search terms *doctor rating*, *physician rating*, *physician-rating*, *physician ranking*, and *quality physicians* produced a list of physician-rating sites currently available in the United States [7,15]. On October 1, 2010, using Google Trends, we identified the most commonly visited physician-rating websites using the number of daily unique visits each website attracted [16,17]. Sites with fewer than 5000 daily unique visits as measured on Google Trends were not included in the analyses. Of note, Google Trends is not an absolute measure of Web traffic. The assumption was that the relative Web traffic volume relationship between different websites was consistent. Websites that had Web traffic that registered on Google Trends but did not allow for user-generated content were not included in the analyses. User-generated content was defined as the ability to rate or comment on the physician.

Rating Content Characteristics of Each Website

We then studied each site to determine the types of rating scales (eg, 1–5, 1–4, 1–100) used and dimensions of care rated (eg, recommend to a friend, waiting room time). All the dimensions of care were identified for each website. To compare different websites, we created a semantic normalization tool. A semantic conversion table was created by first identifying all the different dimensions of care used on each website (Table 1). To facilitate the analysis, each dimension was assigned to 5 categories by three individuals working independently. The 5 different categories were chosen based on the most prevalent rating categories present across various rating websites. There was agreement on 31 of the 35 items, and the group discussed the remaining 4 with the lead author until consensus was reached on the most appropriate category designation: *overall rating, communication skills, access, facilities, and staff*.

Table 1. Semantic conversion table used to normalize different dimensions of care used to rate physicians on the websites

Overall rating	Communication Skills	Access	Facilities	Staff
Overall	Communication	Appointments	Office cleanliness	Courteous staff
Level of trust	Explanation	Approachable	Office setting	Staff
Overall quality of care	Explanation of medications	Doctor availability	Office environment	Staff friendliness
Recommendation	Follow-up	Convenience	Service	Staff helpfulness
Recommend to a friend	Attentive during visit	Ease of appointment	Waiting room	Staff professionalism
Patient satisfaction	Listens and answers questions	Quality of referrals	Facilities	Office friendliness
Likely to recommend	Bedside manner	Make Referrals		
	Helps patient understand	Punctuality		

XSL•FO
RenderX

Analysis of Individual Physician Ratings

Raw data without specific physician identifiers were obtained in October, November, and December 2010 via a nonrandom selection of 4999 online physician ratings from 23 multiple specialties (allergy, cardiology, cardiothoracic surgery, dermatology, endocrinology, gastroenterology, general surgery, hematology, internal medicine, nephrology, neurology, neurosurgery, obstetrics and gynecology, oncology, ophthalmology, orthopedic surgery, otolaryngology, pediatrics, plastic surgery, primary care, pulmonary medicine, rheumatology, and urology) in 25 metropolitan areas (Atlanta, GA; Austin, TX; Baltimore, MD; Boston, MA; Charlotte, NC; Chicago, IL; Colorado Springs, CO; Columbus, OH; Denver, CO; Houston, TX; Los Angeles, CA; Miami, FL; Minneapolis, MN; New Orleans, LA; New York City, NY; Orlando, FL; Phoenix, AZ; Portland, OR; Salt Lake City, UT; San Diego, CA; San Francisco, CA; Raleigh, NC; San Jose, CA; Seattle, WA; and Washington, DC). We chose these cities because they have the highest Internet usage and largest population in the United States [18-20]. The selection of physicians was nonrandom to avoid counting the same physician more than once.

The number of reviews collected from each website varied proportionally by how frequently the websites were visited based on Web traffic estimates from Google Trends. Therefore, the number of reviews from each website was proportional to Web traffic volume assuming that search patterns on Google are similar to those on other search engines.

The sequence of steps followed to acquire each physician rating was to visit the website, enter the city, choose a specialty, enter the largest search radius, and then sort physicians by name when possible. If sorting by name was not possible then location was used. Only reviews that had at least one physician rating completed by a patient within the years 2000–2010 were included in the analyses. Each analyst was assigned a set of metropolitan areas to evaluate physician data.

Cut-offs of 75 (100-point scale), 4 (5-point scale), and 3 (4-point scale) were used to define the favorable threshold for each category of physician-rating website. To compare rankings from different websites with the same rating system, we used a weighted average to accurately represent the overall compiled rating. Only physician-rating sites with the same rating system were compared with one another.

To facilitate analyses, similar dimensions of care—but with different terms used by each website—were grouped into 1 of the 5 categories defined above (*overall rating, access, communication skills, facility, and staff*). For example, wait time, waiting room time, waiting time, and punctuality were all grouped as part of *access* (Table 1).

Results

The Most Commonly Visited Physician-Rating Sites

The 10 most commonly visited online physician-rating websites with user-generated content per Google Trends were HealthGrades.com, Vitals.com, Yelp.com, YP.com, RevolutionHealth.com, RateMD.com, Angieslist.com, Checkbook.org, Kudzu.com, and ZocDoc.com (Table 2).

Table 2. Top 10 most frequently visited physician-rating websites as a relative measure of Web traffic as measured through Google Trends (October-December 2010)

Website	Percentage	Daily unique visits (per Google Trends)
HealthGrades	40%	254,600
Vitals	20%	127,300
Yelp	15%	95,475
Checkbook	7%	44,555
YP	5%	31,825
ZocDoc	4.8%	30,552
AngiesList	3.2%	20,368
RateMD	3%	19,095
RevolutionHealth	1%	6365
Kudzu	1%	6365
Total	100%	636,500

Content Characteristics of Each Website

Patients rated 35 different dimensions of care in the websites, with a median of 4.5 (mean 4.9, SD 2.8, range 1–9) dimensions of care per website (Table 1). There was a varying degree of information available on each physician-rating website. Some websites provide users with information on board certification. Some websites have advertisements and other websites provide users the ability to compare physicians side-by-side. Table 3 summarizes information, features, and the presence of advertisements on physician-rating websites.

XSL•FO
RenderX

Table 3. Information available on the top 10 physician-rating sites

Website	Comments	Board certification	Years in practice	Physician comparison	Advertising	Sanctions
RateMD	Yes	No	Yes	No	Yes	No
Vitals	Yes	Yes	Yes	Yes	Yes	Yes
AngiesList	Yes	No	Yes	No	No	No
HealthGrades	No	Yes	Yes	No	Yes	Yes
YP	Yes	No	No	No	Yes	No
Kudzu	Yes	No	No	No	Yes	No
Yelp	Yes	No	No	No	Yes	No
ZocDoc	Yes	Yes	No	No	No	No
CheckBook	No	Yes	Yes	Yes	No	No
RevolutionHealth	Yes	Yes	Yes	No	Yes	No

Analysis of Individual Physician Ratings

The average rating was 77 (308/400, 77.0%) for sites using a 100-point scale (SD 11, median 76, range 33–100). For sites using a 5-point scale the average rating was 3.84 (76.8%, 2764/3599, SD 0.98, median 4, range 1–5). For sites using a 4-point scale the average was 3.1 (77.5%, 774/1000, SD 0.72, median 3, range 1–4).

The percentage of reviews with a rating of 75 or higher on physician-rating sites with a 100-point scale was 61.5% (246/400). The percentage of reviews with a rating of 4 or higher on sites with a 5-point scale were 57.74% (2078/3599). The percentage of reviews with a rating of 3 or higher on sites with a 4-point scale were 74.0% (740/100) (Table 4 and Figure 1).

Table 4. Physician ratings from the top 10 physician-rating websites with user-generated content. Percentage favorable ratings defined as ≥3 of 4, ≥4 of 5, or ≥75 of 100

Website	Number of reviews evaluated	Percentage of total	Favorable reviews		Overall rating			Lowest rating	Highest rating
			n	%	Mean	SD	Median		
100-Point scales									
Checkbook.org/PatientCentral	350	7%	217	62	77.59	10.48	76.00	34.00	100.00
RevolutionHealth	50	1%	29	57	74.24	16.01	76.00	33.00	100.00
Weighted average	400	8%	246	62	77.17	11.17	76.00	33.00	100.00
5-Point scales									
AngiesList	159	3%	103	65	3.95	0.95	4.00	1.00	5.00
HealthGrades	2000	40%	1139	57	3.82	0.98	4.00	1.00	5.00
Kudzu	49	1%	26	53	3.74	0.96	4.00	1.00	5.00
RateMD	150	3%	87	58	3.84	1.00	4.00	1.00	5.00
Yelp	750	15%	442	59	3.86	0.97	4.00	1.00	5.00
YP	250	5%	158	63	3.93	0.92	4.00	1.00	5.00
ZocDoc	241	5%	123	51	3.77	0.92	4.00	1.00	5.00
Weighted average	3599	72%	2078	58	3.84	0.98	4.00	1.00	5.00
4-Point scale									
Vitals	1000	20%	740	74	3.10	0.72	3.00	1.00	4.00
Total	4999	100%	3064	61.28					

The multiple dimensions of care rated by patients on the physician-rating sites with a 5-point scale had a strong correlation with the overall rating (Pearson correlation, $r = .73$, $P < .001$). In fact, the 20 correlations between each of the 5 dimensions of care measured ranged from .715 to .923 (Pearson correlation, $P < .001$). Even the dimension of care with the lowest correlation coefficient with overall rating (ie, staff rating) was significant: Pearson correlation, $r = .715$, $P < .001$) (Figure 2).

XSL•FO
RenderX

Figure 1. Distribution of percentile ratings for each dimension of care rated on all physician-review sites.

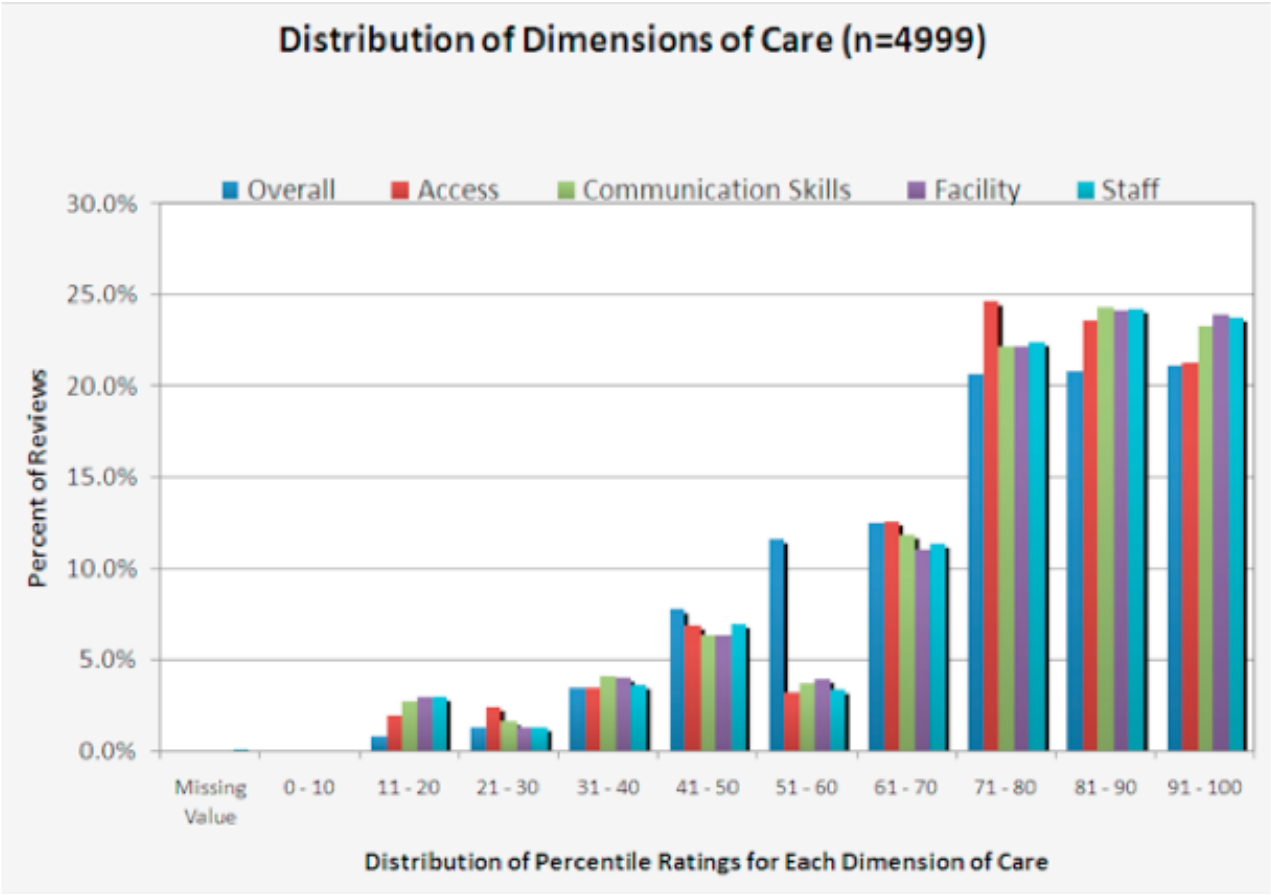
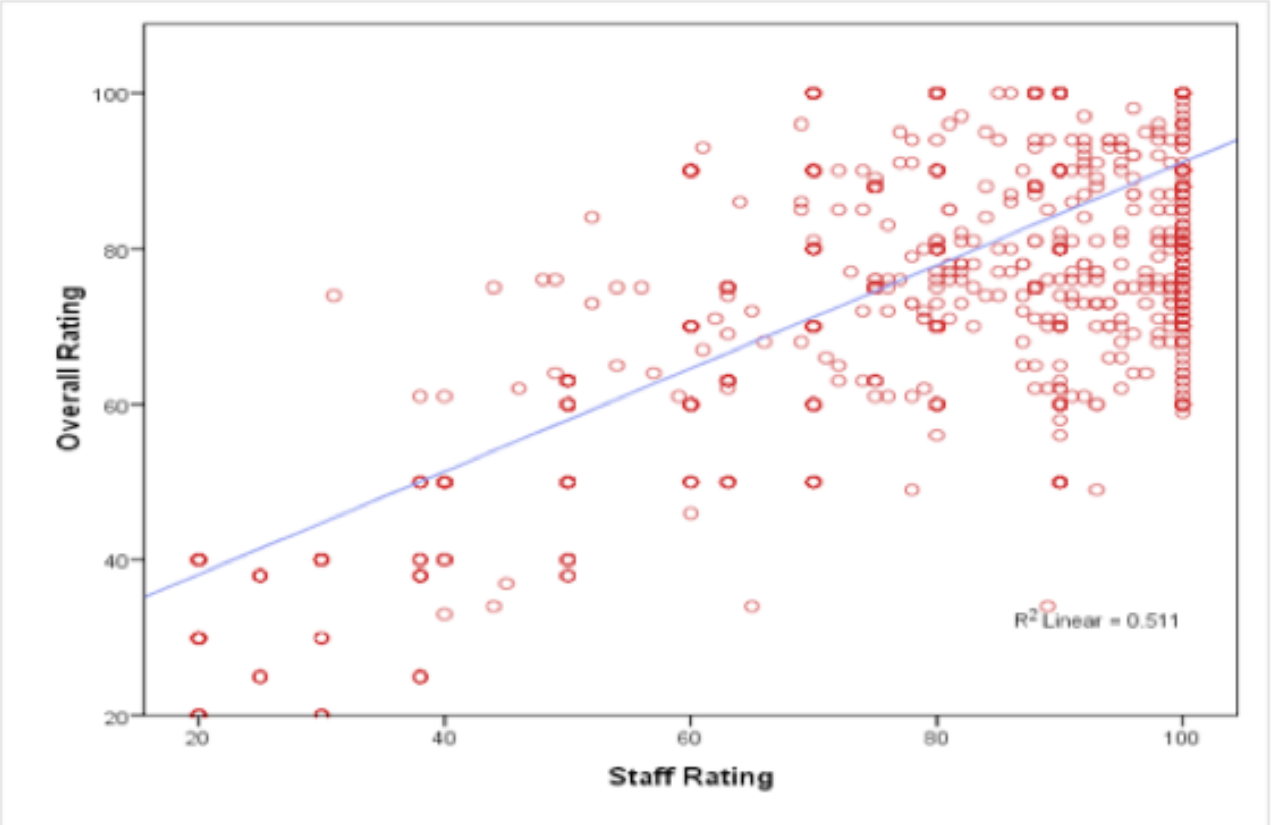


Figure 2. Pearson correlation comparing overall rating versus staff rating (n = 4999, Pearson correlation, $r = .715$, $P < .001$).



XSL•FO
RenderX

Discussion

Results are Consistent with Prior Studies

This analysis of 4999 physician ratings across 10 websites revealed that approximately 2 out of 3 patient reviews are favorable. These results are consistent with a study that found that 88% of 190 reviews of 81 Boston physicians were favorable [15]. In that study, a positive rating was defined as a rating of 3 or 4 in sites with a 4-point scale, or 4 or 5 in sites with a 5-point scale. Our results are also consistent with a report that showed that 67% of all Yelp reviews in 2008 were 4 or 5 stars [21,22]. The majority of physician-rating websites depend on subjective data input and offer limited quantitative information about quality and cost of care. Despite these limitations, patients like these websites because they provide insight into the patient experience from peers [23,24]. This issue is becoming more important, as some physicians and hospitals are caught off guard by online reviews that are critical of their services [8-11]. The optimal content, structure, and rating methods for online physician-rating sites that are most useful deserve further study [1,25-27].

One Feedback Question May be Sufficient to Assess Patient Experience

In all, 35 different dimensions of care were rated by patients in the websites, with an average of 5 questions per site. There was a high correlation between the overall rating of the physician and the other dimensions of care rated (access, communication skills, facility, and staff). This is consistent with using net promoter score methodology to measure customer satisfaction [28]. This raises the issue of whether 1 question may be sufficient to capture the patient's general experience. In fact, the more questions on a rating site, the less likely a patient will complete the survey [29-32]. A single question such as "Would you recommend Dr X to a loved one?" may be as useful as the multitude of specific questions currently surveyed [33]. Also, from the physician's point of view, obtaining actionable information to change communication style, facility, or staff may be better obtained by allowing patients to write in specific feedback and commentary rather than by a scaled survey. In other words, if the facility receives a rating of 1 out of 5 stars, and then the patient comments on how dirty the exam rooms were, then the provider will better understand the low rating.

What makes Physician Ratings Different From Other Professional Service Reviews

Many physicians will take the position that online review sites do not give insight into quality of care. This is valid since obtaining consensus on the definition of quality, even among experts, is challenging. However, patient satisfaction ratings and comments do offer insight into a patient's experience. As more user-generated content is added, the value of ratings will increase. Patient satisfaction is derived from several factors including the baseline expectation of the patient [25,34,35]. Even government agencies, such as the Consumer Assessment of Healthcare Providers and Systems of the Agency for Healthcare Research and Quality and the value-based purchasing programs proposal introduced by the Center for Medicare & Medicaid Services (CMS), are collecting data on the patient experience [36,37]. CMS even launched a portal of their own to allow for physician comparisons [38]. In fact, the German Medical Association assigned the Agency for Quality in Medicine with the task of elaborating quality standards for online physician- and hospital-rating sites [39]. They suggest that a good online rating site defines how the website is financed, separates rating content from advertising, requires user authentication, provides contact information for the site owner, and allows providers to counter offending statements or correct misinformation.

Despite the overall favorable rating of physicians by patients, the topic of physician ratings is rather sensitive [3,6,10,14,40-47]. Advocates for transparency favor a platform that enables patients to truthfully review their experiences. Yet, with further investigation, a few of these "reviews" have become an outlet for patients who are dissatisfied for not getting what they want despite receiving appropriate medical care. Even worse, some reviews are believed to be acts of sabotage from competing providers or organizations [48-50]. Some physicians have even gone as far as getting a court order to remove a review only to find out that such an action invites Internet vigilantes who find it essential that censorship not be tolerated. Also, patient privacy laws make it very challenging to defend against online misinformation and defamation [48-50]. What makes this issue different from other service industries is that "customers" may die or suffer despite appropriate medical care.

Physician-rating websites hosted by insurance companies have been questioned because of the conflict of interest that insurance companies have by reporting data that can potentially drive patients to providers that are cheap and not because they are good [8]. Consumer review organizations have tried though courts to get access to claims data to report volume of care to the public [51]. However, the American Medical Association and US Department of Health Services and Human won an appeal to protect privacy of physician information. Some physicians request their patients to sign agreements that prohibit them from writing about them on physician-rating websites [49,52,53].

Limitations

This study has several limitations. There is an implicit selection bias to websites that depend on the user to actively engage the review site and write a review. In the future, to get more feedback, providers may bundle review requests with online services such as appointments (eg, ZocDoc.com) and social networking sites. This may reduce the selection bias that limits the value of physician ratings. We derived physician-rating site traffic from Google Trends, which is not an absolute measure of total site traffic. Also, the authenticity of the review may be in question [48-50].

XSL•FO
RenderX

Conflicts of Interest

None declared

Multimedia Appendix 1

Video of the presentation of Dr Kadry at the Medicine 2.0 Congress at Stanford University, September 18th, 2011 [http://www.medicine20congress.com/ocs/index.php/med/med2011/paper/view/539] .

[M4V File (M4V Video), 148MB - jmir_v13i4e95_app1.m4v]

References

1. Reimann S, Strech D. The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of English- and German-language sites. BMC Health Serv Res 2010;10:332 [FREE Full text] [doi: 10.1186/1472-6963-10-332] [Medline: 21138579]
2. Taylor H. Harris Interactive. 2010 Aug 4. Press release: "Cyberchondriacs" on the Rise? Those who go online for healthcare information continues to increase URL: http://www.harrisinteractive.com/vault/HI-Harris-Poll-Cyberchondriacs-2010-08-04.pdf [accessed 2011-10-03] [WebCite Cache ID 62AHOpHUj]
3. Segal J. The role of the Internet in doctor performance rating. Pain Physician 2009;12(3):659-664 [FREE Full text] [Medline: 19461833]
4. Keckley PH. Deloitte Center for Health Solutions. 2011. 2011 Survey of Health Care Consumers in the United States: Key Findings, Strategic Implications URL: http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/US_CHS_2011ConsumerSurveyinUS_062111.pdf [accessed 2011-10-02] [WebCite Cache ID 629BcltxM]
5. Fox S, Purcell K. Chronic disease and the Internet. Washington, DC: Pew Internet & American Life Project; 2010 Mar 24. URL: http://www.pewinternet.org/~/media/Files/Reports/2010/PIP_Chronic_Disease.pdf [accessed 2011-10-03] [WebCite Cache ID 62AHTpF0e]
6. Andrews M. Grading doctors online. US News World Rep 2008 Mar 10;144(7):54. [Medline: 18435303]
7. Given R. Vitals.comm and MDx Medical, Inc. 2008 Nov 5. MD Rating Websites: Current State of the Space and Future Prospects URL: http://www.vitals.com/res/pdf/health_care_blog_r.givens_11.5.08.pdf [accessed 2011-10-03] [WebCite Cache ID 62AHhrPb4]
8. Glenn B. The rating game. Patients & insurers are rating the quality of your care. Do you know what they're saying? Med Econ 2008 Dec 5;85(23):18-22. [Medline: 19209532]
9. Mostaghimi A, Crotty BH, Landon BE. The availability and nature of physician information on the internet. J Gen Intern Med 2010 Nov;25(11):1152-1156. [doi: 10.1007/s11606-010-1425-7] [Medline: 20544300]
10. Brayer T. Opposing Views, Inc. 2010 Jul 7. Doctor Sues Online Review Sites for Bad Ratings URL: http://www.opposingviews.com/i/doctor-sues-online-review-sites-for-bad-ratings [accessed 2011-02-22] [WebCite Cache ID 5wh33K0nF]
11. Mills E. CNET News. 2009 Jan 9. Lawsuite Over Yelp Review Settled URL: http://news.cnet.com/8301-1023_3-10139278-93.html [accessed 2011-11-09] [WebCite Cache ID 634TN59ZO]
12. Masnick M. Techdirt. 2010 Nov 18. Doctor Sues Patients Over Bad Yelp Reviews URL: http://www.techdirt.com/articles/20101110/19053611809/doctor-sues-patients-over-bad-yelp-reviews.shtml. [accessed 2011-02-22] [WebCite Cache ID 5wh3dHFjW]
13. Saxenmeyer. Fox Television Stations, Inc. 2010 Nov 16. Chicago Doctor Sues Patients Over Yelp, Citysearch Reviews URL: http://www.myfoxchicago.com/dpp/news/investigative/dr-jay-pensler-yelp-citysearch-reviews-20101115 [accessed 2011-02-22] [WebCite Cache ID 5wh3sycdo]
14. Moyer M. Manipulation of the crowd. Sci Am 2010 Jul;303(1):26, 28. [Medline: 20583659]
15. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946. [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
16. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]
17. Seifter A, Schwarzwalder A, Geis K, Aucott J. The utility of "Google Trends" for epidemiological research: Lyme disease as an example. Geospat Health 2010 May;4(2):135-137 [FREE Full text] [Medline: 20503183]
18. Woyke E. Forbes.com. 2009 Jan 22. America's Most Wired Cities: Seattle Takes the Lead in our Annual List of the Most Broadband-Connected U.S. Cities URL: http://www.forbes.com/2009/01/22/wired-cities-2009-tech-wire-cx_ew_0122wiredcities.html [accessed 2011-02-22] [WebCite Cache ID 5wh42MHIf]
19. Woyke E. Forbes.com. 2010 Mar 2. America's Most Wired Cities: Raleigh Takes the Lead in Our Annual List of the Most Broadband-Connected U.S. Cities URL: http://www.forbes.com/2010/03/02/broadband-wifi-telecom-technology-cio-network-wiredcities_2.html [accessed 2011-02-22] [WebCite Cache ID 5wh49hjsj]
20. American Fact Finder. US Census Bureau. 2010. Search: Population Estimates for the 25 Largest U.S. Cities URL: http://factfinder2.census.gov/faces/nav/jsf/pages/searchresults.xhtml?ref=geo&refresh=t [WebCite Cache ID 62AI7ARkZ]

XSL•FO
RenderX

21. Stoppelman J. Yelp Web Log. 2009. Yelp vs. YouTube: A Ratings Showdown URL: http://officialblog.yelp.com/2009/09/yelp-vs-youtube-a-ratings-showdown.html [accessed 2011-10-03] [WebCite Cache ID 62Agb16XN]
22. Stoppelman J. Bizzy.com. 2011 Apr 4. Yelp Blog Post: Distribution of All Reviews URL: http://blog.bizzy.com/48442434 [accessed 2011-10-03] [WebCite Cache ID 62AhKtzZL]
23. Hay MC, Strathmann C, Lieber E, Wick K, Giesser B. Why patients go online: multiple sclerosis, the internet, and physician-patient communication. Neurologist 2008 Nov;14(6):374-381. [doi: 10.1097/NRL.0b013e31817709bb] [Medline: 19008743]
24. Sciamanna CN, Clark MA, Diaz JA, Newton S. Filling the gaps in physician communication. The role of the Internet among primary care patients. Int J Med Inform 2003 Dec;72(1-3):1-8. [Medline: 14644301]
25. Bleich SN, Ozaltin E, Murray CK. How does satisfaction with the health-care system relate to patient experience? Bull World Health Organ 2009 Apr;87(4):271-278 [FREE Full text] [Medline: 19551235]
26. Petrochuk MA. Leading the patient experience. Driving patient satisfaction and hospital selection. Healthc Exec 2008;23(2):47-48. [Medline: 18481634]
27. Rhodes L, Miles G, Pearson A. Patient subjective experience and satisfaction during the perioperative period in the day surgery setting: a systematic review. Int J Nurs Pract 2006 Aug;12(4):178-192. [doi: 10.1111/j.1440-172X.2006.00575.x] [Medline: 16834578]
28. Kinney WC. A simple and valuable approach for measuring customer satisfaction. Otolaryngol Head Neck Surg 2005 Aug;133(2):169-172. [doi: 10.1016/j.otohns.2005.03.060] [Medline: 16087007]
29. Beebe TJ, Rey E, Ziegenfuss JY, Jenkins S, Lackore K, Talley NJ, et al. Shortening a survey and using alternative forms of prenotification: impact on response rate and quality. BMC Med Res Methodol 2010;10:50 [FREE Full text] [doi: 10.1186/1471-2288-10-50] [Medline: 20529365]
30. Holbrook AL, Krosnick JA, Pfent A. The causes and consequences of response rates in surveys by the news media and government contractor survey research Firms. In: Lepkowski JM, Tucker C, Brick JM, de Leeuw ED, Japec L, Lavrakas PJ, Link MW, Sangster RL, editors. Advances in Telephone Survey Methodology. Hoboken, NJ: John Wiley & Sons; 2008:499-679.
31. Jepson C, Asch DA, Hershey JC, Ubel PA. In a mailed physician survey, questionnaire length had a threshold effect on response rate. J Clin Epidemiol 2005 Jan;58(1):103-105. [doi: 10.1016/j.jclinepi.2004.06.004] [Medline: 15649678]
32. Mailey SK. Increasing your response rate for mail survey data collection. SCI Nurs 2002;19(2):78-79. [Medline: 12510511]
33. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54, 124. [Medline: 14712543]
34. Eisler T, Svensson O, Tengström A, Elmstedt E. Patient expectation and satisfaction in revision total hip arthroplasty. J Arthroplasty 2002 Jun;17(4):457-462. [Medline: 12066276]
35. Kalish GH. Satisfaction with outcome as a function of patient expectation: the national antibiotic patient satisfaction surveys. Health Care Innov 1996;6(5):9-12, 29. [Medline: 10163588]
36. Agency for Healthcare Research and Quality. AHRQ. 2011 Oct 26. Consumer Assessment of Healthcare Providers and Systems (CAHPS) Database URL: https://www.cahps.ahrq.gov/CAHPS-Database/About.aspx [accessed 2011-11-09] [WebCite Cache ID 634TZZl2x]
37. Anonymous. Medicare programs; hospital inpatient value-based purchasing program; proposed rule. Fed Regist 2011;76(9):2454-2491.
38. Medicare. US Department of Health & Human Services. 2011. Physician Compare URL: http://www.medicare.gov/find-a-doctor/provider-search.aspx?AspxAutoDetectCookieSupport=1 [accessed 2011-02-22] [WebCite Cache ID 5wh4Gfs9k]
39. Schaefer C, Schwarz S. [Doctor rating sites: which of them find the best doctors in Germany?]. Z Evid Fortbild Qual Gesundhwes 2010;104(7):572-577. [doi: 10.1016/j.zefq.2010.09.002] [Medline: 21095610]
40. Aungst H. Patients say the darnedest things. You can't stop online ratings, but you can stop fretting about them. Med Econ 2008 Dec 5;85(23):27-29. [Medline: 19209533]
41. Bacon N. Will doctor rating sites improve standards of care? Yes. BMJ 2009;338:b1030. [Medline: 19293223]
42. Gupta S. Rating your doctor. Time 2008 Jan 14;171(2):62. [Medline: 18271202]
43. Hicks A. Positive patient reviews are possible. Fam Pract Manag 2009;16(4):10 [FREE Full text] [Medline: 19626735]
44. Hopkins Tanne J. How patients rate doctors. BMJ 2008;337:a1408. [Medline: 18826978]
45. Hungerford DS. Internet access produces misinformed patients: managing the confusion. Orthopedics 2009 Sep;32(9). [doi: 10.3928/01477447-20090728-04] [Medline: 19751023]
46. Jain S. Googling ourselves--what physicians can learn from online rating sites. N Engl J Med 2010 Jan 7;362(1):6-7 [FREE Full text] [doi: 10.1056/NEJMp0903473] [Medline: 20054044]
47. Pasternak A, Scherger JE. Online reviews of physicians: what are your patients posting about you? Fam Pract Manag 2009;16(3):9-11 [FREE Full text] [Medline: 19492765]
48. Goldberg D. ModernMedicine. 2010 Oct 1. Physicians Have Limited Recourse Against Online Defamation URL: http://www.modernmedicine.com/modernmedicine/Modern+Medicine+Now/Physicians-have-limited-recourse-against-online-de/ArticleStandard/Article/detail/688470 [accessed 2011-02-22] [WebCite Cache ID 5wh4QPJiN]
49. Segal J, Sacopulos MJ, Rivera DJ. Legal remedies for online defamation of physicians. J Leg Med 2009;30(3):349-388. [doi: 10.1080/01947640903146121] [Medline: 19681012]

XSL•FO
RenderX

50. Woodward C. "Anti-defamation" group seeks to tame the rambunctious world of online doctor reviews. CMAJ 2009 May 12;180(10):1010 [FREE Full text] [doi: 10.1503/cmaj.090679] [Medline: 19433818]
51. Anonymous. United States Court of Appeals for the District of Columbia Circuit. 2008 May 1. Consumers' Checkbook v. United States Department of Health and Human Services URL: http://www.ama-assn.org/ama1/pub/upload/mm/395/consumers-checkbook-ama-brief.pdf [accessed 2011-10-03] [WebCite Cache ID 62AISTSue]
52. Boodman SG. The Washington Post. 2009 Jul 21. To Quell Criticism, Some Doctors Require Patients to Sign 'Gag Orders' URL: http://www.usatoday.com/news/health/2009-03-05-doctor-reviews_N.htm [accessed 2011-11-09] [WebCite Cache ID 634dCXWKS]
53. Tanner L. USA Today. 2009 Mar 5. Doctors Seek Gag Orders to Stop Patients' Online Reviews URL: http://www.usatoday.com/news/health/2009-03-05-doctor-reviews_N.htm [accessed 2011-11-10] [WebCite Cache ID 635vMaLwb]

Edited by G Eysenbach; submitted 05.10.11; peer-reviewed by D Strech, J Segal, S Adams; comments to author 25.10.11; revised version received 01.11.11; accepted 07.11.11; published 16.11.11.

Please cite as:
Kadry B, Chu LF, Kadry B, Gammas D, Macario A
Analysis of 4999 Online Physician Ratings Indicates That Most Patients Give Physicians a Favorable Rating
J Med Internet Res 2011;13(4):e95
URL: http://www.jmir.org/2011/4/e95/
doi:10.2196/jmir.1960
PMID:22088924

©Bassam Kadry, Larry F Chu, Bayan Kadry, Danya Gammas, Alex Macario. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 16.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Physician Response Time When Communicating With Patients Over the Internet

Per Egil Kummervold[1,2], PhD; Jan-Are K Johnsen[1,2], PhD

[1]Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway HF, Tromsø, Norway

[2]Department of Clinical Medicine, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway

Corresponding Author:
Per Egil Kummervold, PhD
Norwegian Centre for Integrated Care and Telemedicine
University Hospital of North Norway HF
P.O. Box 6060
Tromsø, 9038
Norway
Phone: 47 07766
Fax: 47 77 75 40 98
Email: per.egil.kummervold@telemed.no

Abstract

Background: Patients want to use electronic communication to access health services more easily. Health authorities in several countries see this as a way to improve health care. Physicians appear to have conflicting opinions regarding the suitability of electronic communication in clinical settings.

Objectives: The aim of our study was to measure how long it actually takes physicians to answer questions from patients through an electronic communication channel, and whether some of the questions are especially time consuming.

Methods: We monitored electronic patient–physician communication. A total of 1113 messages from 14 participating physicians from 7 medical offices were analyzed. The length of questions and answers, and the time physicians spent answering the questions were recorded and analyzed.

Results: Physicians spent an average of 2.3 minutes (median 2 minutes) answering questions from patients. The patients' questions had an average length of 507.1 characters (95% CI 487.4–526.9, SD 336.2), while physicians' answers averaged 119.9 characters (95% CI 189.8–210.0, SD 172.6). The results show that the influence of patient question length on time spent responding was negligible. For the shortest 25% of the questions the answer time was 2.1 minutes (95% CI 1.9–2.3), while it was 2.4 minutes (95% CI 2.2–2.7) for the longest 25%. Even extremely long questions had a minimal impact on the time spent answering them. A threefold increase in question length from patients resulted in only an 18% increase in physician response time.

Conclusions: The study shows the potential clinical usefulness of electronic communication between patients and health care services by demonstrating the potential for saving time.

(J Med Internet Res 2011;13(4):e79) doi:10.2196/jmir.1583

KEYWORDS

Electronic mail; Internet; patients; physicians; patient communication; health communication

Introduction

While the majority of the European population are using the Internet for health purposes, only 1 in 10 Internet users communicate directly with their physician over the Internet. However, this number is rising, increasing from about 5% in 2005 to 9.7% in 2007 [1]. This increase appears to be driven both by patients wanting easier access to health services and by health authorities wanting to make health care more efficient [2].

Among physicians there appear to be conflicting opinions regarding the usefulness of electronic communication in clinical settings. Patt and colleagues [3] reported that some physicians saw email as more convenient, more flexible, and time saving. In contrast, others felt that email could become an added burden, especially if the physician was solely responsible for handling the contact. Also, physicians have expressed concerns that

XSL•FO
RenderX

patients' messages might be inappropriate and inefficient [4]. In sum, physicians' negative perceptions of email contact appear linked to the concern that answering questions from patients will take too much time, and in particular that answering long and complex questions will consume a disproportionate amount of time.

In Norway, purpose-written applications are used for patient–physician communication, since ordinary email does not meet the required security level set by the Norwegian Data Inspectorate. Apart from the user having to log on using a password and one-time codes, the systems provide the same functionality as an email system. From the physicians' point of view, they do, however, integrate more tightly with the electronic patient record. Evidence appears to support that purpose-written applications can be at least as cost effective in large-scale use as email [5]. Also, there is evidence that electronic communication is replacing some traditional inquiries, including visits [6] and telephone calls [7,8], and in general patients hold a positive view of electronic access to health care providers[7].

This study aimed to measure how long it actually takes physicians to answer electronic questions from patients, and whether some of the questions are especially time consuming. Two main hypotheses were posed:

A: The length of questions from patients predicts the time physicians spend answering.

B: The longest questions consume an unreasonably large amount of physicians' time resources.

It is obvious that how long a time a physician uses to compose an answer is correlated with the number of characters he or she is typing. However, the strength of the correlation should be investigated, especially in relation to how long the message from the patient is.

Methods

We asked the 2 suppliers of secure patient communication systems in Norway, Visma Unique [9] and DIPS [10], to provide us with a list of the offices that used the systems actively, and where the systems were integrated with the electronic patient record system. At the time of the study, these 2 systems were the only ones in use in Norway that enabled secure patient–physician communication. From a list of 13 offices, 9 were willing to participate in the study. Due to technical issues, the data from 2 of these offices were inaccessible, leaving us with 7 offices included in the study.

A program logging the time physicians spent answering and the length of the patient questions was installed at the offices included in the study. Time was logged by automatically recording how long the physician took from opening the patient question to sending the answer. In addition, the program recorded the length of the question and of the answer. Prior to sending the answer, the physician was presented with a dialog box indicating the time that had elapsed. This time estimate could then be adjusted if the physician felt this was inaccurate. Both additive and subtractive adjustments could be made. For instance, subtractive adjustments could be made if the physician was interrupted while typing, and additive adjustments could be made if the physician had used more time composing the answer than was recorded by the system. The adjusted time had to be given as an integer. Unadjusted time was therefore also rounded to the closest positive integer, giving a minimum answer time of 1 minute. A total of 380 adjustments were made.

The study ran for 1 year, starting December 2005. A total of 1321 messages were recorded in the period. Physicians sending fewer than 10 messages (n = 1) and physicians not completing the task of returning the data (n = 3) were excluded. Office personnel were not included. This resulted in 14 participating physicians (3 female) and a total of 1113 messages. The physicians had on average worked 15.7 years (range 3–30 years) and had an average patient load of 1441 (range 1100–2300 patients).

The target patient population was all those using primary health services. Earlier studies have shown that young, well-educated persons are overrepresented in using electronic health services [11].

The Regional Committee for Medical and Health Research Ethics approved the study. Hypotheses were investigated by descriptive statistics and linear regression analysis. Data were analyzed using SPSS version 18.0 (IBM Corporation, Somers, NY, USA).

Results

Questions from patients averaged 507.1 (95% CI 487.4–526.9, SD 336.2) characters in length, while the physicians' answers averaged 119.9 (95% CI 189.8–210.0, SD 172.6) characters. Physicians spent an average of 2.3 (SD 2.0) minutes answering questions; 17 (1.5%) of the questions took more than 10 minutes to answer, while 125 (11.2%) of the questions took between 5 and 10 minutes to answer. Table 1 summarizes the descriptive statistics and Table 2 shows the time the participating physicians spent answering patient questions.

Table 1. Descriptive statistics of question length for patients and physicians

	Minimum	Maximum	Mean	SD
Patient question length (number of characters)	100	3315	507.1	336.2
Physician answer length (number of characters)	14	1634	119.9	172.6

XSL•FO
RenderX

Table 2. Response time of participating physicians

Physician ID	Number of questions answered	Mean (minutes)	Median (minutes)		
			25%	50%	75%
A	68	2.7	1	2	3
B	187	2.1	1	2	2
C	123	3.2	1	2	5
D	24	2.5	2	2	3
E	18	1.5	1	1	2
F	20	1.8	1	1	1
G	74	3.0	2	3	4
H	46	2.8	1	2	3
I	39	1.2	1	1	1
J	12	1.2	1	1	1
K	79	1.7	1	1	2
L	218	2.1	1	1	3
M	82	1.8	1	1	2
N	123	2.0	1	1	2
Total	1113	2.3	1	2	3

We expected that the length of the patients' questions would predict response time (hypothesis A). The hypothesis was investigated through regression analyses. Two models were tested. The first model included only the length of the patients' questions. While the model significantly explained variance (P = 0.007), the effect size was small (beta = .08) and the overall fit of the model was very low ($R2$ = .01). The second model included also the length of the physicians' answers (Table 3) and showed better fit ($R2$ = .26). The results indicate that the influence of patient question length on response time is negligible (beta = –.05, P = .05) compared with the length of the physician's answer (beta = .53, P < .001).

Table 3. Summary of regression analysis for patient question length and physician answer length predicting response time (minutes)

Model	B	SE	Beta	t	P value
(Constant)	1.20	.10		12.02	<.001
Patient question length	.00	.00	–.05	–1.94	.05
Physician answer length	.01	.00	.53	19.75	<.001

Patient questions were categorized based on their length. Patient questions were divided into quartiles each containing 25% of the messages (Table 4). This confirmed that for most of the questions, the effect of question length on answer time was negligible. The answer time was 2.1 minutes (95% CI 1.9–2.3) for the shortest 25% of questions and 2.4 minutes (95% CI 2.2–2.7) for the longest 25%.

Table 4. Time physicians spent answering patients' questions by question length

Patient question length		Number of answers	Physician answer time in minutes (95% CI)
Quartile	Number of characters		
1	0–308	278	2.1 (1.9–2.3)
2	309–398	280	2.3 (2.1–2.5)
3	399–594	277	2.2 (1.9–2.5)
4	595–3315	278	2.4 (2.2–2.7)

Hypothesis B states that the longest questions would consume an unreasonably large amount of physicians' time resources. These questions were defined as being the top 10% of questions (110 questions) with regard to length (>916 characters) (see Table 5).

XSL•FO
RenderX

Table 5. Time physicians spent answering the longest 10% of patients' questions

	Number of questions	Patient question (mean number of characters)	Physician answer time (minutes)
Shortest 90% of questions (≤916 characters)	1003	420	2.2
Longest 10% of questions (>916 characters)	110	1300	2.6

As shown in Table 5, the 10% longest patient questions were approximately 3 times the length of shorter questions (420 characters versus 1300 characters). However, the physicians spent on average 18% more time answering the 10% longest questions (2.6 minutes compared with 2.2 minutes). These results were not in favor of hypothesis B.

Discussion

The results give mixed support to the hypotheses. As expected, the length of patients' questions predicted answer time, but the analysis also shows that the predictive value is negligible compared with the length of the physicians' answers. We did not find conclusive support for the hypothesis that very long patient messages should have a large effect on physician answer length and answering time. Instead, we observed a modest increase in physician answer length and only a marginal increase in answering time related to extremely long patient questions.

As noted, one of the main reasons physicians are skeptical about electronic communication is the potential for increased workload [12]; for instance, physicians might fear that patients would overuse it or that responding to questions would be time consuming. The results of the current study show, however, that these specific concerns might be unfounded. While it does take extra time to read long questions from the patient, this does not have large effects on the total time used by physicians to answer patient inquiries. In fact, a threefold (300%) increase in patient question length resulted in only an 18% increase in physician response time. Even though the average numbers may support the effectiveness of an electronic communication system, and other studies indicate that responses to email messages do not take more time than responses to nonelectronic patient messages [13], one may still question whether extreme cases will jeopardize these effects in a real-life office setting. Based on the current results these concerns appear unfounded. The average physician response time to a patient message was 2.3 minutes using the systems described in this study. Only 1.5% of questions took more than 10 minutes to answer. When compared against the average consultation time in Norwegian general practice (15–20 minutes) [14], even these unusual cases will have to be regarded as time saving, if the electronic messages substitute patients' office visits [6]. It is, however, unlikely that electronic messages can be a substitute for office visits in a one-to-one relationship. Other studies have shown that electronic messages can replace phone calls [7,8], and it is very likely that electronic messaging will find relevance as a supplement to personal encounters, for instance by recommendation of ethical guidelines [15].

Results from other investigations indicate that patients are willing to adapt to guidelines regarding the focus and content of messages [4], which should help to minimize the potential problem of lengthy patient questions. Obviously, the time-saving potential is highly dependent on electronic messages substituting for patient office visits [6].

Limitations

This study included a considerable proportion of Norwegian physicians using electronic patient communication at the time the study was performed. It is not self-evident that the result would be valid for all physicians using similar services. An alternative approach would be to select a random sample of all physicians using electronic communication. At the time of the study, only a few Norwegian physicians were offering electronic communication services. A random sample could therefore be biased toward physicians being positive to electronic communication. In some countries, for instance Denmark [1], it has become mandatory for physicians to offer electronic communication services. In such contexts, a similar study based on random selection would be feasible.

The current study does not involve analysis of the content of the messages. The main challenge in doing this would be that it would require written consent from every patient. However, a prior study in similar populations has shown that only a small proportion of these messages are used for simple administrative purposes such as scheduling [11]. Instead, the majority of the patient messages are concerned with health-related questions, and requesting prescriptions, test results, and documentation for medical leave.

The average time spent answering messages might be influenced by factors such as workload and reimbursement policies. This limits the external validity of the current results. However, the investigated relationships between variables (eg, that the length of patients' messages had limited impact on the answer time), rather than their absolute values, are much more likely to also be valid in other cultural contexts.

Conclusions

Studies have demonstrated how email and electronic messaging systems can be used to promote balanced and patient-centered communication [16], in support of clinical decision making [17-19]. We believe the results reported here further extend the clinical usefulness of electronic communication between patients and health care providers by demonstrating the potential for saving time.

XSL•FO
RenderX

Acknowledgments

This study was supported by a grant from Helse Nord RHF. Thanks to Gro Berntsen, Trine S Bergmo, and Ellen Ersfjord for various contributions to this project.

Conflicts of Interest

None declared

References

1. Kummervold PE, Chronaki CE, Lausen B, Prokosch HU, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
2. Wallwiener M, Wallwiener CW, Kansy JK, Seeger H, Rajab TK. Impact of electronic messaging on the patient-physician interaction. J Telemed Telecare 2009;15(5):243-250. [doi: 10.1258/jtt.2009.090111] [Medline: 19590030]
3. Patt MR, Houston TK, Jenckes MW, Sands DZ, Ford DE. Doctors who are using e-mail with their patients: a qualitative exploration. J Med Internet Res 2003;5(2):e9 [FREE Full text] [doi: 10.2196/jmir.5.2.e9] [Medline: 12857665]
4. White CB, Moyer CA, Stern DT, Katz SJ. A content analysis of e-mail communication between patients and their providers: patients get the message. J Am Med Inform Assoc 2004;11(4):260-267 [FREE Full text] [doi: 10.1197/jamia.M1445] [Medline: 15064295]
5. Caffery L, Smith AC, Scuffham PA. An economic analysis of email-based telemedicine: a cost minimisation study of two service models. BMC Health Serv Res 2008;8:107 [FREE Full text] [doi: 10.1186/1472-6963-8-107] [Medline: 18495045]
6. Bergmo TS, Kummervold PE, Gammon D, Dahl LB. Electronic patient-provider communication: will it offset office visits and telephone consultations in primary care? Int J Med Inform 2005 Sep;74(9):705-710. [doi: 10.1016/j.ijmedinf.2005.06.002] [Medline: 16095961]
7. Liederman EM, Lee JC, Baquero VH, Seites PG. Patient-physician web messaging. The impact on message volume and satisfaction. J Gen Intern Med 2005 Jan;20(1):52-57. [doi: 10.1111/j.1525-1497.2005.40009.x] [Medline: 15693928]
8. Lin CT, Wittevrongel L, Moore L, Beaty BL, Ross SE. An Internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
9. Minhelse.no. Visma Unique AS. 2009. Online Kontakt med Helsepersonell URL: https://www.minhelse.no/login/ [accessed 2011-08-25] [WebCite Cache ID 61CuvtTe0]
10. DIPS ASA, Profdoc Norge AS. 2011 Jan 18. minDoktor.no URL: https://www.mindoktor.no/publicservice/index.jsp [accessed 2011-08-25] [WebCite Cache ID 61Cuxev47]
11. Kummervold PE, Trondsen M, Andreassen H, Gammon D, Hjortdahl P. [Patient-physician interaction over the internet]. Tidsskr Nor Laegeforen 2004 Oct 21;124(20):2633-2636 [FREE Full text] [Medline: 15534640]
12. Kittler AF, Carlson GL, Harris C, Lippincott M, Pizziferri L, Volk LA, et al. Primary care physician attitudes towards using a secure web-based portal designed to facilitate electronic communication with patients. Inform Prim Care 2004;12(3):129-138. [Medline: 15606985]
13. Leong SL, Gingrich D, Lewis PR, Mauger DT, George JH. Enhancing doctor-patient communication using email: a pilot study. J Am Board Fam Pract 2005;18(3):180-188 [FREE Full text] [Medline: 15879565]
14. Hjortdahl P, Hunskår S. [1.2 Health problems and the use of health services in general practice]. 3rd edition. In: Hunskår S, editor. [General Practice]. Oslo, Norway: Gyldendal Norsk Forlag; 2006:28-34.
15. Bovi AM, Council on Ethical and Judicial Affairs of the American Medical Association. Ethical guidelines for use of electronic mail between patients and physicians. Am J Bioeth 2003;3(3):W-IF2. [doi: 10.1162/152651603322874771] [Medline: 14735881]
16. Roter DL, Larson S, Sands DZ, Ford DE, Houston T. Can e-mail messages between patients and physicians be patient-centered? Health Commun 2008;23(1):80-86. [doi: 10.1080/10410230701807295] [Medline: 18443995]
17. Perlis TE, Des Jarlais DC, Friedman SR, Arasteh K, Turner CF. Audio-computerized self-interviewing versus face-to-face interviewing for research data collection at drug abuse treatment programs. Addiction 2004 Jul;99(7):885-896. [doi: 10.1111/j.1360-0443.2004.00740.x] [Medline: 15200584]
18. Tideman RL, Chen MY, Pitts MK, Ginige S, Slaney M, Fairley CK. A randomised controlled trial comparing computer-assisted with face-to-face sexual history taking in a clinical setting. Sex Transm Infect 2007 Feb;83(1):52-56. [doi: 10.1136/sti.2006.020776] [Medline: 17098771]
19. Wolford G, Rosenberg SD, Rosenberg HJ, Swartz MS, Butterfield MI, Swanson JW, et al. A clinical trial comparing interviewer and computer-assisted assessment among clients with severe mental illness. Psychiatr Serv 2008 Jul;59(7):769-775 [FREE Full text] [doi: 10.1176/appi.ps.59.7.769] [Medline: 18586994]

XSL•FO
RenderX

Edited by G Eysenbach; submitted 13.06.10; peer-reviewed by S Ross, V Della Mea, M Koivunen, N Atehortua; comments to author 10.11.10; revised version received 17.06.11; accepted 24.06.11; published 01.11.11.

Please cite as:
Kummervold PE, Johnsen JAK
Physician Response Time When Communicating With Patients Over the Internet
J Med Internet Res 2011;13(4):e79
URL: http://www.jmir.org/2011/4/e79/
doi:10.2196/jmir.1583
PMID:22044909

©Per Egil Kummervold, Jan-Are K Johnsen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 01.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Public Access and Use of Health Research: An Exploratory Study of the National Institutes of Health (NIH) Public Access Policy Using Interviews and Surveys of Health Personnel

Jamie O'Keeffe[1], MEd; John Willinsky[1], PhD; Lauren Maggio[2], MS(LIS)

[1]Stanford University School of Education, Stanford, CA, United States

[2]Lane Medical Library, Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
Jamie O'Keeffe, MEd
Stanford University School of Education
485 Lasuen Mall
Stanford, CA, CA, 94305
United States
Phone: 1 650 862 9093
Fax: 1 650 725 7412
Email: jamie.okeeffe@stanford.edu

Abstract

Background: In 2008, the National Institutes of Health (NIH) Public Access Policy mandated open access for publications resulting from NIH funding (following a 12-month embargo). The large increase in access to research that will take place in the years to come has potential implications for evidence-based practice (EBP) and lifelong learning for health personnel.

Objective: This study assesses health personnel's current use of research to establish whether grounds exist for expecting, preparing for, and further measuring the impact of the NIH Public Access Policy on health care quality and outcomes in light of time constraints and existing information resources.

Methods: In all, 14 interviews and 90 surveys of health personnel were conducted at a community-based clinic and an independent teaching hospital in 2010. Health personnel were asked about the research sources they consulted and the frequency with which they consulted these sources, as well as motivation and search strategies used to locate articles, perceived level of access to research, and knowledge of the NIH Public Access Policy.

Results: In terms of current access to health information, 65% (57/88) of the health personnel reported being satisfied, while 32% (28/88) reported feeling underserved. Among the sources health personnel reported that they relied upon and consulted weekly, 83% (73/88) reported turning to colleagues, 77% (67/87) reported using synthesized information resources (eg, UpToDate and Cochrane Systematic Reviews), while 32% (28/88) reported that they consulted primary research literature. The dominant resources health personnel consulted when actively searching for health information were Google and Wikipedia, while 27% (24/89) reported using PubMed weekly. The most prevalent reason given for accessing research on a weekly basis, reported by 35% (31/88) of survey respondents, was to help a specific patient, while 31% (26/84) were motivated by general interest in research.

Conclusions: The results provide grounds for expecting the NIH Public Access Policy to have a positive impact on EBP and health care more generally given that between a quarter and a third of participants in this study (1) frequently accessed research literature, (2) expressed an interest in having greater access, and (3) were aware of the policy and expect it to have an impact on their accessing research literature in the future. Results also indicate the value of promoting a greater awareness of the NIH policy, providing training and education in the location and use of the literature, and continuing improvements in the organization of biomedical research for health personnel use.

(J Med Internet Res 2011;13(4):e97) doi:10.2196/jmir.1827

KEYWORDS

Health policy; evidence-based practice; information storage and retrieval; access to information; information literacy; health personnel

XSL•FO
RenderX

Introduction

> *Funded research should be accessible and open without cost (or with very minimal cost) to all providers…[since charging] such high rates for access to crucial information is detrimental to health care.*

The above quote was submitted by a health care provider in response to our recent survey on awareness of the 2008 National Institutes of Health (NIH) Public Access Policy [1]. The policy requires all research publications resulting from NIH funding to be made publicly accessible through deposit in the National Library of Medicine's PubMed Central within 12 months of publication. This NIH policy is expected to provide public access to some 80,000 biomedical research articles annually [2] and represents a broader trend within scholarly communication toward "open access" [3]. For example, a number of funding agencies, journals, and institutions, such as the Wellcome Trust, *Journal of Medical Internet Research*, and Stanford University, have adopted open access policies. It is imperative that these policies be assessed for their (potential) impact and ability to guide future policies.

Health personnel's access to research literature has taken on greater cogency with the ongoing emphasis on evidence-based practice (EBP) in health care. EBP is the "conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" [4]. *Best evidence* is defined as the current evidence from the research literature [4,5]. However, research on information seeking among health personnel has repeatedly demonstrated their bias toward easily accessible information [6-8]. For example, in a study of 47 physicians during a half-day of practice, Covell [9] found that physicians raised a total of 269 questions about patient management. Roughly two-thirds of these questions were not pursued, a finding that has been repeated by Gorman and Helfand [10]. Of the 30% of questions that were pursued, the physicians most frequently sought answers from other health professionals [9].

Other studies, such as Haug's [11] meta-review of 12 physician studies, found that physicians prefer to consult local and easily accessible information sources, such as colleagues or local textbooks. Although nurses have been less frequently studied, they seem to have similar preferences for local, easily accessible information sources [6]. Slawson and Shaughnessy [8] call this phenomenon "satisficing," in which health personnel are "satisfied with the information they have at hand, sacrificing quality for convenience." Instead of seeking the best evidence, many providers rely on summaries and guidelines, whether or not they are evidence-based [8].

These findings raise important questions about health personnel's level of access to "best evidence" from the research literature. Currently, the leading sources of easily accessible evidence are point-of-care (POC) services, such as UpToDate, ClinicalEvidence, and DynaMEd, among others, which generally provide synthesized accounts of evidence on major medical topics. In the study reported here, all participants had access through their institution to UpToDate, which is self-described as a clinical decision support system based on current evidence [12]. While the availability of such resources raises questions about the added value that increased access to the research literature provides, studies have demonstrated the limitations of preappraised sources in answering complex clinical questions when compared with the research literature [13,14]. While health personnel have been found to prefer POC resources to PubMed for primary literature [15], POC resources have been questioned on the level [16,17] and currency [18] of evidence that they utilize in creating their syntheses of medical topics [16].

This study measured current research usage among health personnel from a hospital and community clinic in order to assess the potential impact of the NIH Public Access Policy as well as provide a preliminary set of measures for future policy assessments. These measures include (1) perceived quality of current access to information, (2) source and frequency of access, (3) search strategies, (4) reasons for access, and (5) awareness of the NIH Public Access Policy.

Methods

Population

The Stanford University Institutional Review Board reviewed and approved the study in June 2010. The study was performed at two health care facilities in Northern California, an independent teaching hospital and a community clinic, from July 2010 through November 2010. The teaching hospital has fewer than 200 beds. The community clinic includes multiple clinical sites involving fewer than 100 physicians and nurses. In 2009, the vast majority of the clinic's patients were at or below the federal poverty level, and almost all were either uninsured or on public health insurance (eg, Medi-Cal).

Interviews

During July and August of 2010, 14 interviews were conducted. Interview participants were recruited via each institution's email listserv. Of those interviewed, 14 agreed to participate, 4 of whom also completed an optional PubMed interview and training. In all, 7 participants were hospital employees and 3 were clinic employees; 6 women and 4 men were interviewed ranging in age from 30 to 50 years. The participants included 5 physicians, 4 nurses, and 1 neuropsychologist; 2 of the physicians also held administrative roles. Interview prompts focused on participant work roles, tasks, characteristics of information needs, sources of information, and outcomes. Questions were directed toward information use, in general (eg, "What resources do you use to get the information you need?"), and use of research, in particular (eg, "Do you ever consult research articles in journals, either in print or online?")

Surveys

The survey was informed by previous studies [19] and the preliminary results of the interviews. Survey respondents were recruited via each institution's email listserv. Members of each institution's listserv were sent a recruitment email that contained a link to the online survey and was open to any participant that accessed the link. The survey itself and the ten questions within it were voluntary. The online survey contained a total of five screens: the first screen included background information about

XSL•FO
RenderX

the study, screens two through four each contained two questions, and screen five contained four questions.

Survey respondents included 90 health personnel across the two sites, 88 of which completed the survey. Of the survey respondents, 32% (28/88) were from the clinic, representing a 32% response rate for clinic personnel. A total of 68% (60/88) of the respondents were from the hospital; however, a response rate for this site could not be determined as the number of health personnel on the email listserv was unknown. Of the respondents, 79% (69/87) were female and 21% (18/87) were male. In addition, 46% (40/88) were physicians, 26% (23/88) were nurses, and 28% (25/88) were other health personnel, such as physician assistants, psychologists, and social workers.

Results

Quality of Access to Health Information

Survey respondents were asked to characterize their current level of access to research articles as excellent, good, poor, very poor, or not applicable. Of the survey respondents, 65% (57/88) reported their level of access to be good or excellent, while 32% (28/88) respondents reported access to be poor or very poor (see Table 1). A physician at the hospital summarized the issue of a lack of access:

> *I've looked [online] at some abstracts for articles that I couldn't get. I've thought, 'Wow, that's a really good article. It's too bad I can't read it'*. [H-P10 (Bracketed numbers refer to interview participants)]

Table 1. Perceived quality of access

	n (%) (Total N = 88)
Excellent: I have access to all the research articles that I need.	8 (9%)
Good: I have access to most of the research articles that I need.	49 (56%)
Poor: I frequently have difficulty getting the research articles that I need.	26 (30%)
Very poor: I always have great difficulty getting the research articles I need.	2 (2%)
Not applicable: I do not need access to research articles.	3 (3%)

Source and Frequency of Access

Survey respondents were asked how often they consult different types of information sources (see Table 2). UpToDate was identified apart from other POC services (such as MDConsult) because it alone was universally available to all participants. In that light, the results for UpToDate, as a POC, can be combined with the category *other POC* services, suggesting that 77% (67/87) of participants used a POC on a weekly basis.

Table 2. Source and frequency of access

	Weekly n (%)	Monthly n (%)	Less Frequently n (%)	Totals N (%)
Other medical professionals	73 (83%)	7 (8%)	8 (9%)	88 (100%)
Reference books or websites	59 (69%)	18 (21%)	10 (10%)	86 (100%)
Clinical guidelines/protocols	41 (47%)	26 (30%)	21 (23%)	88 (100%)
UpToDate	39 (45%)	4 (5%)	44 (50%)	87 (100%)
Review articles	34 (39%)	30 (34%)	24 (27%)	88 (100%)
Original research articles	28 (32%)	29 (33%)	29 (35%)	88 (100%)
Other POCs (eg, MDConsult)	28 (32%)	13 (15%)	46 (53%)	87 (100%)
Personal journal subscriptions	21 (24%)	19 (22%)	48 (54%)	88 (100%)

Search Strategies

Respondents and interviewees used a number of strategies to find health information outside of POC resources. In all, 67% (57/85) of respondents reported weekly use of Google or Wikipedia (as a guide to related sources), and 27% (24/89) reported weekly use of PubMed or MEDLINE (see Table 3). Only 9% (8/87) of survey respondents reported weekly access to an online university library collection.

XSL•FO
RenderX

Table 3. Search strategy

	Weekly n (%)	Monthly n (%)	Less Frequently n (%)	Totals N (%)
Google or Wikipedia	57 (67%)	18 (21%)	10 (12%)	85 (100%)
PubMed or MEDLINE	24 (27%)	21 (24%)	44 (50%)	89 (100%)
Remote access to university library	8 (9%)	1 (1%)	78 (90%)	87 (100%)

Reasons for Access

In all, eight themes emerged from our interview data related to reasons for accessing research (see Table 4). Of those eight, the three most frequently reported reasons for consulting research on a weekly basis were "informing my understanding of a specific patient" (35%, 31/88), "out of general interest" (31%, 26/88), and "informing and updating clinical practice, in general" (25%, 22/88).

Table 4. Reasons for accessing health research

	Weekly n (%)	Monthly n (%)	Less Frequently n (%)	Totals N (%)
With regard to a specific patient	31 (35%)	23 (26%)	34 (39%)	88 (100%)
Out of general interest	26 (31%)	18 (21%)	40 (48%)	84 (100%)
Informing clinical practice in general	22 (25%)	29 (33%)	37 (42%)	88 (100%)
Educating patients and their families	15 (17%)	25 (29%)	47 (54%)	87 (100%)
Training or informing health personnel	16 (18%)	17 (19%)	55 (63%)	88 (100%)
Preparing for school or licensure	11 (13%)	7 (8%)	70 (89%)	88 (100%)
Preparing for presentation or teaching	7 (8%)	16 (18%)	65 (74%)	87 (100%)
Writing protocols, articles, or books	3 (3%)	8 (9%)	77 (88%)	88 (100%)

Policy Awareness

Survey respondents were asked whether they were previously aware of the NIH Public Access Policy, and 27% (23/86) of participants reported being aware, while 73% (63/86) reported they were not familiar with the policy. The survey respondents reporting awareness of the policy also indicated more frequently utilizing research (65% on a weekly basis) than respondents reporting no policy awareness. When asked whether free online access enabled by the policy is likely to have an impact on the frequency of their accessing research articles, 83% (72/87) of respondents reported that some impact is likely, 12% (10/87) reported that no impact is likely, and 6% (5/87) did not know if it would have an impact.

Discussion

Quality of Access to Health Information

When asked about the quality of their current access to research articles, 65% (57/88) of participants in this study responded that it was good or excellent, while 32% (28/88) of respondents said it was poor or very poor. These responses need to be taken in light of the universal access among this sample to UpToDate. One survey respondent added that he receives complimentary journal subscriptions and, if he "had to pay, [he] wouldn't be reading them."

The interview results yielded a similar picture of access to research, with six of the ten interviewees describing their quality of access as good or excellent, three as poor or very poor, and one as not applicable. Among those who felt that their access was inadequate, one hospital-based physician described the consequences of a low level of access to research:

> *It's a pretty regular basis where people are having problems [accessing articles]. If you can't pull up the articles, it's difficult to get all the information you need…it impacts our patient care.* [H-P06]

A nurse at the hospital commented on the frequency of this issue, saying, "nine out of ten times, I can't get all the articles that I want" [H-N09]. Current literature reports that perhaps a quarter of articles in PubMed are publicly accessible [20-22]. One might draw from this that potentially a third of the health personnel in this study, who feel their current access is inadequate, would be likely to welcome and take advantage of the NIH Public Access Policy.

Source and Frequency of Access

In the survey, 83% (73/88) of respondents reported that they consult their colleagues on a weekly basis, which aligns with previous research [23]. Reference books and websites were utilized weekly by 69% (59/86) of the respondents. However, close to half (45%, 39/87) used UpToDate weekly. Additionally, the use of other POC resources such as MDConsult and Clinical Resources at Ovid suggest that the majority regularly relied on POC products. As previous studies have shown [24] and as our participant interviews affirmed, UpToDate is a preferred information source. A physician at the community clinic said:

> *UpToDate is pretty much what most clinicians use when they have a clinical question. Most people just*

XSL•FO
RenderX

> *read the article, and they say, 'This is what UpToDate says, so this is what I'm doing'.* [C-P01]

She also described UpToDate's appeal: "[It's] simple. It's the iPod model" [C-P01]. A second physician at the clinic made similar observations but also discussed the limitations of UpToDate:

> *It's from one center....The protocols [are] very predigested.…You know, it's usually just written by a bunch of experts sitting in a room. Yeah, they have all the access to evidence, but you never know the sources of bias.* [C-P02]

To our knowledge, few studies have assessed the quality of medical information in POC products such as UpToDate. Of those that do, one recent study compared the volume of content, editorial quality, and evidence-based methodology of 18 POC products [17]. UpToDate ranked in the top quartile for editorial quality and evidence-based methodology and the high intermediate quartile for the volume of content. However, another recent study that did an in-depth bibliometric analysis of five popular POC products (UpToDate, ACP PIER, Clinical Evidence, DynaMed, FirstCONSULT) found that three of the five products, including UpToDate, had nearly 50% or more citations from 2001 or earlier for the four topics analyzed (hypertension, asthma, carbon monoxide poisoning, and hyperlipidemia) [18]. The authors also found surprisingly little overlap in citations among the POC products for the four medical topics analyzed. While the authors acknowledge that this study only examined summaries of four topics, they conclude that it "reveals surprising and critical information about these POC products: they can vary greatly in content, from the raw number of citations, to the types of evidence, to the currency of those citations" [18]. They advise users to "judiciously appraise POC product information content" when using this information for EBP [18].

Our study suggests that fewer participants read review articles or research articles than turn to this POC. This finding must be considered in relation to the amount of current research available, the current difficulties in searching the biomedical literature, and how these factors will gradually improve over time. Given the potential limitations of "evidence" in POC products as well as in certain clinical guidelines [25-28], it may be necessary for health personnel to have access to the primary literature for the best evidence necessary for EBP.

In the interviews, a hospital physician described the frequency of his need for access: "It's pretty much on a day-to-day basis that you're looking for something" [H-P06]. A physician at the community clinic described her general information needs as "constant" [C-P01]. She attributed this to her natural curiosity but also to the complexity of patient issues and to the lack of easily accessible consultations from specialists.

> *I have to look up stuff all the time, mainly because I'm curious about things, but also because people ask me questions, and I don't always know the answers. The level of questions…tend to be very sophisticated…because patients are so complex, [and] because our providers are really adept and are left to manage a lot of things that, in other places, specialists would be taking care of.* [C-P01]

These finding are consistent with previous studies [9,11] that have found colleagues and reference textbooks to be physicians' preferred information sources [6]. Several studies [29,30] have also cited research journals as "a primary mechanism for continuing medical education" [31]. Studies on nurses report a similar preference for local sources of information including patient and lab data as well as colleagues [32-35]. Nurses may also consult nursing journals if the content is relevant to the patient issue [11].

Search Strategies

Although the majority of respondents turned to POC services, many also reached outside of these resources, turning to Google or Wikipedia more frequently than PubMed. This preference may be due to an interest in background information, which these resources provide, whereas research articles contain more specific foreground information. Studies have shown that 50% to 70% of physicians use Wikipedia as an information source [36]. Empirical studies have begun to emerge on the quality of Wikipedia's medical content. These studies find few factual errors but also a general lack of depth and ease of understanding [37].

Ease of access and searching of these tools may contribute to their relative popularity. For example, a physician and administrator at the clinic said that she uses Google because it is "easy" and "in front of me" [C-P03]. She said that she typically sticks to "Google, UpToDate, or a consultant" because of this ease of access [C-P03].

Some health personnel used Google to access relevant research. As a physician at the clinic described: "In this setting, since you would have to pay for the articles…if you're real interested in the answer to your question, you can just do some Google searches and shotgun around and see if some institutional setting has an unprotected link to it" [C-P02]. Similarly, a nurse at the hospital said that she always starts with Google but is careful to "always look for websites that have .edu or. org or .gov" as opposed to "commercial websites" [H-N09].

PubMed or MEDLINE was the fourth most frequently reported resource, reported to be used on a weekly basis by 27% (24/89) of health personnel. The interview participants elaborated on the various uses of PubMed. A physician at the clinic said that one of her "most common PubMed searches," which she conducts approximately weekly, is searching for "a review article from core clinical journals within the last five years" [C-P01]. A hospital physician said he uses PubMed "to educate other people and myself," to prepare for talks and for rare patient cases, such as "an unusual hemoglobinopathy" (H-P10). He added that he typically uses PubMed "in retrospect" because "it takes a while if you're researching articles to find something" [H-P10].

Other participants described their PubMed strategies and frustrations. A hospital nurse described how she typically accesses PubMed via Google: "I just enter what I want to know on Google…and if I see PubMed, I always go to that because I trust it as a source of the latest research" [H-N09]. She added,

XSL•FO
RenderX

however, that she has "not had much success navigating [PubMed]" [H-N09]. A medical fellow at the hospital described how he "used to use PubMed" but switched to Google Scholar because it seemed "more user friendly," "pulled things up better," and generally "works better" [H-P06]. A hospital nurse added, "I primarily use [PubMed] if I know the exact citation" [H-N05]. Training in PubMed has been shown to foster favorable information-seeking behavior, such as increased searches in MEDLINE [38]. Our findings suggest that training in PubMed may bolster the impact of the NIH Public Access Policy.

Comparisons of preappraised information sources with the primary literature have shown, for example, that preappraised sources could not provide answers to 40% of the complex clinical questions, 95% of which had previously been answered by the primary research literature [13]. Other studies have similarly pointed to the value of access to the research literature in addition to POCs [14,39].

Only 9% (8/87) of survey respondents had access to a medical school library's online resources. This affiliation provides the user with remote access via a secure user name and password to the library's online full-text journal subscriptions. It should be noted that individuals with these privileges acknowledge upon receipt of their access that they will not share this information, as sharing is a violation of the library's contract with publishers. Related to the value of a password, a physician at the community clinic noted that remote access is one of the major benefits of training residents: "You either don't get paid, or don't get paid much, but you *do* get access to that library" [C-P02]. Several interviewees described how individuals with passwords to the medical school library are often asked to retrieve articles online for others. A hospital nurse, who frequently used this method of access, said:

> *I never [pay for articles]. I just call different friends that have access to different services and do what I need to do.* [H-N09]

A hospital physician and clinical professor described how "people will ask [him] to get articles through the library" as often as "a couple times a month" [H-P10]. He concluded, "My assumption is that they're searching, and for whatever reason, it's not an open-access article, and so they're asking me to get it from [the university library]" [H-P10]. A medical fellow at the hospital said that he has been using his friend's password to access the university library for a number of years. When asked whether he could remember any instances of having difficulty accessing a particular article, he said:

> *I've been cheating the system for so long that I don't remember....Pretty much every single article is on their site.* [H-P06]

Reasons for Access

In the current study, the most popular reason for consulting research on a weekly basis was for "informing my understanding of a specific patient" (35%, 31/88), a finding that has been demonstrated in previous research [40]. Several interviewees related stories of patient care that hinged on access to research articles. For example, a medical fellow at the hospital described how access to research was necessary in treating a complex case:

> *There was a kid who had hemorrhagic cystitis…and we were trying to figure out an appropriate way to treat that, and it's a rare, complicated thing. And so having access to the articles gave potential treatment options, which, in this case, was actually injecting the bladder with an agent that helps you to clot. Within a day, we got the articles that we needed, and we were able to start that treatment, which helped the kid.* [H-P06]

In this case, the medical fellow used his friend's password to a university library to gain remote access to the necessary articles. He said that the full-text articles were necessary because with abstracts only "you're not going to get enough information to make a meaningful decision" (H-P06). A community clinic physician described a similar scenario involving patient treatment, in which he was unable to freely access the article that he needed:

> *A question came up a little while ago about what oral antibiotics you can take for bone infections in the outpatient setting…it made the difference between the patient getting six months of IV antibiotics at home…or just getting six weeks' oral antibiotics without needing a intravenous capillary…and I ended up paying 50 bucks for the article.* [C-P02]

The second most frequently reported reason for accessing research on a weekly basis was out of general interest (31%, 26/84). As a clinic physician said, "At any given moment, I'll have two folders full of things that I'm reading" [C-P01]. Similarly, a nurse at the hospital said that she reads anything that interests her: "It kind of intrigues [me] - I'm naturally curious - to read about head lice, which has no relevance to anything I do clinically, but I think it's interesting" [H-N05].

Studies have also demonstrated the association between physician use of an online evidence system with patient admissions, suggesting that evidence use was related to patient care [41]. Additional research is needed to empirically test whether access itself would lead to increased use of research literature. A physician at the community clinic nicely summarized the issue at hand:

> *I view [reading research] as necessary, but something that I do very little, just to be completely honest…. You just learn that you don't have that many resources available to you, or, if you do, it's a pain to get them at the point of care. So it's the question of the chicken or the egg. Do we not do it because we're not that interested? Or do we not do it because it's such a pain to find it, and we become accustomed to not doing it?* [C-P02]

The physician raises an important empirical question that warrants further study. He seemed to be describing a sort of *learned helplessness* that occurs when the needed information is not available. Future research under consideration, given the warrant provided by the results of this study, is a randomized controlled trial (RCT) comparing the uses of research by a

XSL•FO
RenderX

high-access group that receives relatively complete access to the biomedical research literature and another *same-access group* with typical (unchanged) access. A third group, a *high-access group with support from medical librarians*, could also be included in order to examine the value-added benefit of training and support. Ideally, this research would be longitudinal, as few studies examine health personnel use of online evidence over an extended period of time [42].

Limitations of the Study

While encouraging, our study has limitations. For example, participation for health personnel was only solicited from two health care sites, which means that these findings may or may not be applicable outside of these two institutions. Also, this study only analyzed self-reported behaviors and preferences, which may differ from actual practices. As Covell [9] has demonstrated, physicians tend to overestimate their actual use of printed information resources, while underestimating their use of peer consultation. Lastly, a relatively small number of health personnel were surveyed and, as with many voluntary surveys, our sample may be biased. Future plans to increase sample size and the number of participants will be considered for additional research.

Finally, future research may benefit from examining the potential impact of the NIH policy among other participant groups, such as patients or researchers. Since many individuals are seeking health information online - attracted to the convenience, coverage, and anonymity of online health information [43] - it may be beneficial to track the potential impact of the NIH policy within this population. Similarly, the NIH Public Access Policy is expected to be beneficial for biomedical researchers without access to well-funded research libraries or who do not work in one of the hundred-plus of the world's poorest countries that qualify for the Health Access to Research program (HINARI) [44]. Of course, a recent RCT demonstrated that articles assigned to the open access condition received more downloads than control articles, and the authors concluded that the true beneficiaries of open access publishing may be consumers, not producers, of the medical literature [45], such as the health personnel studied here.

Conclusions

This study establishes a preliminary measure of current research use, interest, and barriers among a sample of health personnel in hospital and community clinic contexts. While health personnel have limited time available for consulting additional sources and are already equipped with POC services, the results still provide grounds for expecting the NIH Public Access Policy to have a positive impact on EBP and health care more generally, given that between a quarter and a third of the participants (1) frequently access research literature, (2) express an interest in having greater access, and (3) are aware of the NIH policy and expect it to have an impact on their accessing the research literature. Additional measures are warranted if health personnel and their patients are going to maximize the benefits from this increased access to research through policy promotion, medical education, continuing website improvements to PubMed, and research on the nature of clinical practices and decision making in light of this increased access to the research literature.

Acknowledgments

This study was funded by the Science of Science and Innovation Policy program of the National Science Foundation, award number SMA-1038188.

Conflicts of Interest

None declared

References

1. National Institutes of Health Public Access Policy. NIH Public Access Policy Details. Washington, DC: US Department of Health and Human Services; 2008. URL: http://publicaccess.nih.gov/policy.htm [accessed 2011-11-03] [WebCite Cache ID 62vO48RCv]
2. Zerhouni EA. NIH's Public Access Policy: Subcommittee on Courts, the Internet, and Intellectual Property. 2008 Presented at: 145th Meeting of the National Cancer Advisory Board; Feb 5, 2008; Washington, DC.
3. Willinsky J. The Access Principle: The Case for Open Access to Research and Scholarship. Cambridge, Mass.: MIT Press; 2006.
4. 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]
5. Mayer D. Essential Evidence-Based Medicine. New York, NY: Cambridge University Press; 2004.
6. Case D. Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior. Amsterdam, Netherlands: Elsevier/Academic Press; 2007.
7. Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the "evidence cart". JAMA 1998 Oct 21;280(15):1336-1338 [FREE Full text] [Medline: 9794314]
8. Slawson DC, Shaughnessy AF. Teaching evidence-based medicine: should we be teaching information management instead? Acad Med 2005 Jul;80(7):685-689. [Medline: 15980087]

XSL•FO
RenderX

9. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? Ann Intern Med 1985 Oct;103(4):596-599. [Medline: 4037559]
10. Gorman PN, Helfand M. Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered. Med Decis Making 1995;15(2):113-119. [Medline: 7783571]
11. Haug JD. Physicians' preferences for information sources: a meta-analytic study. Bull Med Libr Assoc 1997 Jul;85(3):223-232. [Medline: 9285121]
12. UpToDate. Waltham, MA: UpToDate, Inc; 2011. URL: http://www.uptodate.com/index [accessed 2011-03-25] [WebCite Cache ID 5xTnicVSn]
13. Koonce TY, Giuse NB, Todd P. Evidence-based databases versus primary medical literature: an in-house investigation on their optimal use. J Med Libr Assoc 2004 Oct;92(4):407-411. [Medline: 15494755]
14. Patel MR, Schardt CM, Sanders LL, Keitz SA. Randomized trial for answers to clinical questions: evaluating a pre-appraised versus a MEDLINE search protocol. J Med Libr Assoc 2006 Oct;94(4):382-387. [Medline: 17082828]
15. Hoogendam A, Stalenhoef AF, Robbé PF, Overbeke AJ. Analysis of queries sent to PubMed at the point of care: observation of search behaviour in a medical teaching hospital. BMC Med Inform Decis Mak 2008;8:42 [FREE Full text] [doi: 10.1186/1472-6947-8-42] [Medline: 18816391]
16. Farrell A. An evaluation of the five most used evidence based bedside information tools in Canadian health libraries. Evidence Based Library and Information Practice 2008;3(2):3-17 [FREE Full text]
17. Banzi R, Liberati A, Moschetti I, Tagliabue L, Moja L. A review of online evidence-based practice point-of-care information summary providers. J Med Internet Res 2010;12(3):e26 [FREE Full text] [doi: 10.2196/jmir.1288] [Medline: 20610379]
18. Ketchum AM, Saleh AA, Jeong K. Type of evidence behind point-of-care clinical information products: a bibliometric analysis. J Med Internet Res 2011;13(1):e21 [FREE Full text] [doi: 10.2196/jmir.1539] [Medline: 21335319]
19. Mullaly-Quijas P, Ward DH, Woelfl N. Using focus groups to discover health professionals' information needs: a regional marketing study. Bull Med Libr Assoc 1994 Jul;82(3):305-311. [Medline: 7920341]
20. Matsubayashi M, Kurata K, Sakai Y, Morioka T, Kato S, Mine S, et al. Status of open access in the biomedical field in 2005. J Med Libr Assoc 2009 Jan;97(1):4-11. [doi: 10.3163/1536-5050.97.1.002] [Medline: 19159007]
21. McVeigh ME, Pringle JK. Open access to the medical literature: how much content is available in published journals? Serials 2005 Mar;18(1):45-50. [doi: 10.1629/1845]
22. Björk BC, Welling P, Laakso M, Majlender P, Hedlund T, Gudnason G. Open access to the scientific journal literature: situation 2009. PLoS One 2010;5(6):e11273 [FREE Full text] [doi: 10.1371/journal.pone.0011273] [Medline: 20585653]
23. McCord G, Smucker WD, Selius BA, Hannan S, Davidson E, Schrop SL, et al. Answering questions at the point of care: do residents practice EBM or manage information sources? Acad Med 2007 Mar;82(3):298-303. [doi: 10.1097/ACM.0b013e3180307fed] [Medline: 17327723]
24. McKibbon KA, Fridsma DB. Effectiveness of clinician-selected electronic information resources for answering primary care physicians' information needs. J Am Med Inform Assoc 2006;13(6):653-659 [FREE Full text] [doi: 10.1197/jamia.M2087] [Medline: 16929042]
25. Abramson J, Starfield B. The effect of conflict of interest on biomedical research and clinical practice guidelines: can we trust the evidence in evidence-based medicine? J Am Board Fam Pract 2005;18(5):414-418 [FREE Full text] [Medline: 16148253]
26. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Q 2004;82(4):661-687. [doi: 10.1111/j.0887-378X.2004.00327.x] [Medline: 15595946]
27. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365(9453):82-93. [doi: 10.1016/S0140-6736(04)17670-8] [Medline: 15639683]
28. Steinberg EP, Luce BR. Evidence based? Caveat emptor!. Health Aff (Millwood) 2005;24(1):80-92. [doi: 10.1377/hlthaff.24.1.80] [Medline: 15647218]
29. Currie BF. Continuing education from medical periodicals. J Med Educ 1976 May;51(5):420. [Medline: 1263234]
30. Curry L, Putnam RW. Continuing medical education in Maritime Canada: the methods physicians use, would prefer and find most effective. Can Med Assoc J 1981 Mar 1;124(5):563-566. [Medline: 7471001]
31. Marshall JG. Issues in clinical information delivery. Library Trends 1993;42(1):83-107.
32. Corcoran-Perry S, Graves J. Supplemental-information-seeking behavior of cardiovascular nurses. Res Nurs Health 1990 Apr;13(2):119-127. [Medline: 2320757]
33. Wakeham M. The information seeking behaviour of nurses in the UK. Information Services and Use 1992;12(2):131-140.
34. Blythe J, Royle JA. Assessing nurses' information needs in the work environment. Bull Med Libr Assoc 1993 Oct;81(4):433-435. [Medline: 8251980]
35. Pettirgrew KE. Lay information provision in community settings: how community health nurses disseminate human services information to the elderly. Library Quarterly 2000;70(1):47-85.
36. Comer B. Medical Marketing & Media. New York, NY: MM&M/Haymarket Media, Inc; 2009 Apr 21. Docs look to Wikipedia for condition info: Manhattan research URL: http://www.mmm-online.com/docs-look-to-wikipedia-for-condition-info-manhattan-research/article/131038/ [accessed 2011-03-31] [WebCite Cache ID 5xbOSjwZk]

XSL•FO
RenderX

37. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]
38. Byrnes JA, Kulick TA, Schwartz DG. Information-seeking behavior changes in community-based teaching practices. J Med Libr Assoc 2004 Jul;92(3):334-340. [Medline: 15243639]
39. Schilling LM, Steiner JF, Lundahl K, Anderson RJ. Residents' patient-specific clinical questions: opportunities for evidence-based learning. Acad Med 2005 Jan;80(1):51-56. [Medline: 15618093]
40. Younger P. Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. Health Info Libr J 2010 Mar;27(1):2-10. [doi: 10.1111/j.1471-1842.2010.00883.x] [Medline: 20402799]
41. Westbrook JI, Coiera EW, Gosling AS. Do online information retrieval systems help experienced clinicians answer clinical questions? J Am Med Inform Assoc 2005;12(3):315-321 [FREE Full text] [doi: 10.1197/jamia.M1717] [Medline: 15684126]
42. Magrabi F, Westbrook JI, Kidd MR, Day RO, Coiera E. Long-term patterns of online evidence retrieval use in general practice: a 12-month study. J Med Internet Res 2008;10(1):e6 [FREE Full text] [doi: 10.2196/jmir.974] [Medline: 18353750]
43. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
44. Parker K. Bridging the partnership divide: HINARI over eight years. 2009 Presented at: The Tenth International Congress on Medical Librarianship; Aug 31 - Sept 4 2009; Brisbane, Australia.
45. Davis PM. Open access, readership, citations: a randomized controlled trial of scientific journal publishing. FASEB J 2011 Jul;25(7):2129-2134. [doi: 10.1096/fj.11-183988] [Medline: 21450907]

Abbreviations

EBP: evidence-based practice
POC: point of care
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 17.04.11; peer-reviewed by J Till; comments to author 14.06.11; revised version received 25.06.11; accepted 07.07.11; published 21.11.11.

Please cite as:
O'Keeffe J, Willinsky J, Maggio L
Public Access and Use of Health Research: An Exploratory Study of the National Institutes of Health (NIH) Public Access Policy Using Interviews and Surveys of Health Personnel
J Med Internet Res 2011;13(4):e97
URL: http://www.jmir.org/2011/4/e97/
doi:10.2196/jmir.1827
PMID:22106169

©Jamie O'Keeffe, John Willinsky, Lauren Maggio. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 21.11.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Study of Innovative Features in Scholarly Open Access Journals

Bo-Christer Björk[1], PhD

Hanken School of Economics, Helsinki, Finland

Corresponding Author:
Bo-Christer Björk, PhD
Hanken School of Economics
P.O. Box 479
Helsinki, 00101
Finland
Phone: 358 50 3553425
Fax: 358 9 43133275
Email: Bo-Christer.Bjork@hanken.fi

Abstract

Background: The emergence of the Internet has triggered tremendous changes in the publication of scientific peer-reviewed journals. Today, journals are usually available in parallel electronic versions, but the way the peer-review process works, the look of articles and journals, and the rigid and slow publication schedules have remained largely unchanged, at least for the vast majority of subscription-based journals. Those publishing firms and scholarly publishers who have chosen the more radical option of open access (OA), in which the content of journals is freely accessible to anybody with Internet connectivity, have had a much bigger degree of freedom to experiment with innovations.

Objective: The objective was to study how open access journals have experimented with innovations concerning ways of organizing the peer review, the format of journals and articles, new interactive and media formats, and novel publishing revenue models.

Methods: The features of 24 open access journals were studied. The journals were chosen in a nonrandom manner from the approximately 7000 existing OA journals based on available information about interesting journals and include both representative cases and highly innovative outlier cases.

Results: Most early OA journals in the 1990s were founded by individual scholars and used a business model based on voluntary work close in spirit to open-source development of software. In the next wave, many long-established journals, in particular society journals and journals from regions such as Latin America, made their articles OA when they started publishing parallel electronic versions. From about 2002 on, newly founded professional OA publishing firms using article-processing charges to fund their operations have emerged. Over the years, there have been several experiments with new forms of peer review, media enhancements, and the inclusion of structured data sets with articles. In recent years, the growth of OA publishing has also been facilitated by the availability of open-source software for journal publishing.

Conclusions: The case studies illustrate how a new technology and a business model enabled by new technology can be harnessed to find new innovative ways for the organization and content of scholarly publishing. Several recent launches of OA journals by major subscription publishers demonstrate that OA is rapidly gaining acceptance as a sustainable alternative to subscription-based scholarly publishing.

(J Med Internet Res 2011;13(4):e115) doi:10.2196/jmir.1802

KEYWORDS

Scholarly publishing; open access; Internet; peer review

Introduction

Development of Scientific Journal Publishing

The scientific journal as an institution dates back to the late 17th century. Until the Second World War, scholarly journals were mainly published by scientific societies, and subscriptions were primarily individual and often linked to society membership. After 1950, the number of journals increased rapidly, and commercial publishers entered the market to meet the increased demand for outlets. Today there are almost 30,000 peer-reviewed scholarly journals indexed in Ulrich's periodicals directory, and there are several thousand journals more,

XSL•FO
RenderX

particularly journals published in languages other than English. Approximately 1.5 million articles are published yearly [1].

From the perspective of the scientific community as a whole, the scholarly journal fulfills a number of functions [2]. Gierveld, for instance, names four such functions: current awareness, archival recording, priority claim, and quality control [3]. Over the years, the format of scientific articles hasn't changed much. In most disciplines, articles are a few pages long and are bundled into regularly appearing issues that are collected into yearly volumes. Citations provide the "glue" that links the articles into the context of a scientific field's body of knowledge. The quality assurance mechanisms have undergone a gradual change as the anonymous peer-review process evolved into an "industry standard" in the 20th century [4].

During the past two decades, the scientific journal publishing process has undergone more change than during the preceding three centuries together. Today, almost all major subscription journals are available in both paper and electronic formats, giving academics at major universities instant access to thousands of journals, mainly via bundled e-licensing agreements with the major publishers. Publishers have also adopted Web-based manuscript and review management systems. From the readers' viewpoint, a very significant improvement is the emergence of general Web search engines, as well as specialized services dedicated to scientific literature (ie, Google Scholar), which facilitate discovering and tracking publications enormously.

The scientific publishing industry has a peculiar oligopolistic structure, which has created extremely high barriers for new entrants and has ensured the major publishers a high level of profitability [5]. Because of this situation, mainstream publishers have had little incentive to experiment with the radical innovation of open access (OA), and part of the potential of the Web for dissemination of scientific knowledge has remained untapped. Instead, individual scholars and new start-up publishers have taken the initiative and have in a short time launched several thousand OA journals.

Open Access

The fundamental principle of open access, that the results of science should be openly accessible to anybody, is perfectly in harmony with the fundamental ethos of science and also with the interests of authors, academic institutions, and research funders. Open access can be achieved in two ways: via direct electronic OA publishing (ie, gold OA) or, alternatively, by publication in traditional subscription journals combined with parallel posting of the manuscript openly on the Web (ie, green OA) [6]. This paper deals only with direct OA. There are several comprehensive studies of both routes to OA [7-9].

A recent study has estimated that the number of OA journals increased by 500% and the number of articles by 900% during the decade 2000-2009 [10]. The difference between the two growth measures is explained by the fact that the average yearly number of articles published per OA journal rose from around 20 to 40 during the period. In 2009, there were around 4800 active OA journals, which published approximately 190,000 articles. An estimated 7.7% of all peer-reviewed articles were published in full OA journals [10].

Behind these aggregate numbers, the population of OA journals is very heterogeneous in size, funding mechanism, Web features, and the method of peer review and scientific quality. The academics and publishers behind these journals have experimented with many of the parameters of scholarly journal publishing, sometimes successfully, sometimes ending in failure. So far, there have mainly been reports about individual OA journals [11-15] focusing on the features and experiences gained from the journal in question. In this study, we attempt to paint a picture of the broader spectrum of these innovations and to draw some tentative conclusions as to where scholarly OA publishing is moving.

Innovation and Scholarly Publishing

Innovations typically occur in transition periods when technical inventions such as the printing press, steam power, electricity, or the Internet radically change the production conditions and cost structures of whole industries, enabling entrepreneurs to start offering new products or services.

There is a rich literature on the concept of innovation. Tidd et al [16] discuss the "4Ps" of innovation, from a company's business model perspective. The 4Ps are: (1) product innovation, that is, changes in the products/services which an organization offers; (2) process innovation, that is, changes in the ways in which these products/services are created and delivered; (3) position innovation, that is, changes in the context in which the products/services are introduced; and (4) paradigm innovation, that is, changes in the underlying mental models that frame what the organization does.

Of these categories, product, process, and paradigm innovation are easily applicable to our context. Although a peer-reviewed journal article in its traditional printed format can be seen as a product, it is more useful to view the publication of scholarly journal articles as service provision, since the product is not consumed when read and the key issue is disseminating the information as efficiently as possible to potential readers. The process perspective is also important since both electronic publishing and open access publishing enable major changes in the process [17,18]. The paradigm innovation in the context of publishing are the ideas of making the journals available to the whole world for free and of funding the service by means other than charging the readers.

Baregheh et al [19] define innovation as "the multistage process whereby organizations transform ideas into improved products, services, or processes in order to advance, compete, and differentiate themselves successfully in their marketplace." This definition stresses that the driving force of innovation is to improve the competitive position of market players, which leads to the concept of a business model. For our purposes, a useful definition is: "The essence of a business model is that it defines the manner by which the business enterprise delivers value to customers, entices customers to pay for value, and converts those payments to profit: it thus reflects management's hypothesis about what customers want, how they want it, and how an enterprise can organize to best meet those needs, get

XSL•FO
RenderX

paid for doing so, and make a profit" [20]. Since many of the publishers of scholarly journals are scientific societies, groups of independent scholars, and so on, this definition should be extended from a commercial profit motive alone to also ensuring the long-term economic sustainability of the publishing operations.

A key element in this definition is the concept of *value to customers*. A scientific journal can only be successful and sustainable if it succeeds in delivering value appreciated by its customers and in covering the costs of its operations by monetary income or voluntary efforts. In the open access context, the authors are the key customers. One could argue that the editors and reviewers should be included as customers as well, or perhaps more appropriately, as partners. Editors and reviewers make very significant contributions, in particular, they contribute to journals with rigorous peer review in exchange for the personal network and the prestige they gain within their academic communities by being associated with the journal and are hence receiving a sort of service from the journal in exchange for their value-added work. The success of any start-up peer-reviewed journal is very much dependent on attracting this type of contributor.

The effects of the Web on scholarly publishing can be seen as consisting partly of effects enabled by the e-infrastructure as such and those enabled by opening up the e-versions with no access restrictions. In an industry with high barriers to the entry of new companies, established journal publishers, in the first instance, have striven to use the medium of the Internet to enhance the current business model. The more radical innovation of opening up the Web versions of journals has forced new journals and publishers to come up with alternative business models at the same time as it has offered the chance of offering a different type of service to authors. This innovation can be compared with radical changes in other information-related industries, as exemplified by successful companies and community services like Skype, Wikipedia, and Red Hat.

A useful two-dimensional framework for discussing the development in scholarly publishing during the last 15 years can be constructed using the principles of dissemination technology and access (Table 1).

Table 1. A typology of scientific peer-reviewed journals

	Restricted Access	Open Access
Paper only	Traditional printed journal	-
Paper and electronic	All major publishers today	Immediate or delayed OA to electronic version
Electronic only	Very rare type of journal	Full OA journal

The position of a journal in this framework defines many of the border conditions for the features a journal can experiment with. Until the emergence of the Internet, paper printing was the only option, and, in that mode, restricted access for buyers and subscribers was the only viable alternative. (For a brief period, CD-ROM, which can only function in the restricted access mode was also tried, but this was more common for conference proceedings).

Paper and electronic is the dominant solution today, as almost all major publishers have launched parallel electronic versions of their journals. Publishers of journals with parallel electronic versions have, in general, restricted access to the e-versions, which has facilitated two new types of distribution mechanisms: the bundled e-licenses, with sometimes over a thousand titles, and the e-commerce, with individual articles on a pay-per-view basis. Since the electronic versions are usually just copies of the articles in the print issues, the structure of journals with a fixed number of regular issues has usually been retained. Many publishers nowadays post accepted and processed papers on the electronic journal sites well in advance of the actual publishing in order to speed up the dissemination, which is otherwise slowed down by articles queuing in line for a fixed number of yearly issues. Other features the electronic medium has made possible are citation linking (ie, Crossref) and alerting emails that contain tables of content and other notifications. Most journals nowadays use electronic manuscript handling systems (either proprietary or open-source), which facilitate the peer-review process without changing the process or the end product, only making the process more efficient.

Some publishers, in particular professional society publishers, have opened up access to the e-versions, which can be accessed for free, directly, or with a delay. This free access is subsidized by income from the print versions or from subscribers wishing immediate access. This is in line with the fundamental purpose of such societies, which is the efficient knowledge dissemination in their subject area. Societies can also see the offering of free e-versions as a way to attract new members and of branding themselves.

The restricted access electronic-only journal is still quite rare. It is well suited for newly founded high volume journals, for instance, which include data sets or case reports.

The last option is the full-fledged open access journal, most of which were "born OA," which has more freedom than journals in the other categories, with the exception of the revenue model, where readers cannot be charged. So far, electronic-only OA journals have been published mainly by individual academics or start-up OA publishing companies, which tend to use article-processing charges to fund their operations.

The starting point of our further discussion of innovations in scholarly journal publication is the realization that delivering value to the author is what primarily matters for the success of a journal. The collection of article processing charges is only possible if the authors (and, in increasing cases, their funders) perceive that they get value for their money. And this value is in turn dependent on the type of service the journal provides including how widely articles are read and also the branding the journal offers in terms of prestige for the author. Central

XSL•FO
RenderX

services that authors are seeking, and which provide them with value, include the prestige of being published in a highly regarded journal, the assurance of being widely read by the relevant readership, speed of publication, and high likelihood of acceptance [21-23].

What kind of new features, then, can open access journals offer that haven't been possible in traditional journals? Textbox 1 contains a non-exhaustive list of some features.

Some of these features can also be present in subscription-based journals, although they are more commonly found in OA journals.

Textbox 1. Features of Open Access Journals That Differ From Traditional Journals

Paradigm:

- The universal accessibility per se

Process:

- Cost savings by the use of volunteers for tasks other than peer review
- Cost savings by the use of open-source software
- Cost savings by the use of third party e-portals

Revenue:

- Funding by article processing charges

Product/service:

- Broader or narrower journal topics due to the global reach
- Novel peer review methods
- Faster article publication cycles
- More flexibility in the layout and structure of articles
- Interactivity for after-publication discussions
- Easy reusability of the (digital) content

Methods

Choosing 20 to 25 journals randomly (or even using a stratified random sample) from the 5000 journals in the DAOJ at the time the study was started would probably not have yielded a very interesting set of journals to study. The vast majority of OA journals consist of individually created journals published by academics, universities, or scientific societies and typically do not use article processing charges for funding [24]. They also tend to use traditional peer-review methods and the articles look much like paper ones. The major innovation is thus the open accessibility itself, not further innovations made possible by the combination of OA and electronic delivery.

A different strategy is to choose key or outlier cases, which have characteristics making them either highly representative or atypical. In theory, the websites of all open access journals listed in the Directory of Open Access Journals (DOAJ) [25] could have been visited and the journals could have been classified according to a predefined list of features. Based on this search, interesting cases could have been identified in a systematic fashion. The list of features would probably have grown from an initial one, as new features would have emerged during the search. For all practical purposes, such a search would have been extremely resource demanding and was ruled out from the start.

Instead, the case journals were identified based on a literature search of articles and conference presentations about open access and also based on the author's previous extensive knowledge of OA publishing and his personal network. The aim was both to find highly representative cases (where one case is used to represent a large number of journals with fairly similar characteristics) and to find rare atypical cases where journals have experimented with new features. Some journals published in languages other than English were included. The cases also span different revenue models and different sizes ranging from a few articles per year to thousands. The process of case selection was also iterative in the sense that additional interesting candidates came up during conversations with stakeholders or during the study of already selected cases.

In an earlier study, our research group proposed a periodization of the development of OA journals into three periods: a pioneering stage from 1993 through 1999, an innovation period from 2000 through 2004, and a consolidation period beginning in 2005 [10]. This periodization partly influenced the choice of case journals so that each period was represented by several journals. Some basic data about the case journals is shown in Table 2.

For each included journal, information was found using secondary sources (ranging from blog discussions, conference presentation material, and general newspaper items to articles published in peer-reviewed journals), by studying the journal

XSL•FO
RenderX

website itself, and by using journal indexes (ie, from the Institute for Scientific Information [ISI] journal citation reports).

There were two options for how the following narrative could have been structured: by journal or by innovative feature. The first option seemed more natural since it would enable the context around the journals to be presented first. In the second option, the case journals would be loosely grouped under a number of dominant themes. The early journals tend to be presented first with the latest newcomers towards the end, but the order is not rigorously followed. The conclusions section is, on the other hand, structured according to the innovations discussed.

Table 2. The journals discussed in this paper listed according to the start year of OA publishing.

Journal	Year OA Began	Type of Journal	Type of Publisher	APC (USD)	Impact Factor	Information Technology (IT) Platform	Number of Articles in 2010[a]
Elore	1994	Born OA	Society		-	OJS	9
Journal of Electronic Publishing	1995	Born OA	University	-	-	Own	20
Information Research	1995	Born OA	Scholar	-	0.4	OJS	32
Medical Education Online	1996	Born OA	Commercial	800	-	OJS	21
Electronic Transactions on Artificial Intelligence	1997	Born OA	Scholar	-	-	Own	ceased
The International Journal of Design Computing	1997	Born OA	Scholar	-	-	Own	ceased
British Medical Journal	1998	E-version OA	Society	-	13.6	Publisher's	~1300
Journal of Medical Internet Research	1999	Born OA	Scholar	1990	4.7	Own (OJS fork)	64
Malaria Journal	2002	Born OA	Commercial	1775	2.9	Publisher's	360
Journal of Negative Results in Biomedicine.	2002	Born OA	Commercial	1665	-	Publisher's	9
PLoS Biology	2003	Born OA	Non-Commercial	2900	12.9	Publisher's	~200
BMC Medicine	2003	Born OA	Commercial	2265	3.9	Publisher's	60
Hydrology and Earth System Sciences	2004	Born OA	Commercial	per page	2.4	Publisher's	~300
Advances in Difference Equations	2004	Born OA	Commercial	600	0.8	Publisher's	133
Nucleic Acids Research	2005	Converted to OA	Univ. Press	2770	7.4	Publisher's	~1200
African Journal of Food, Agriculture, Nutrition and Development	2005	E-version OA	Society	-	-	Bioline	111
PLoS ONE	2006	Born OA	Non-Commercial	1350	4.3	Publisher's	> 7000
Diagnostic Pathology	2006	Born OA	Commercial	1670	1.4	Publisher's	83
Open Medicine	2007	Born OA	Scholar	1235	-	OJS	24
Boletim do Museu Paraense Emílio Goeldi	2007	E-version OA	Society	-	-	Scielo	27
Open Information Science Journal	2008	Born OA	Commercial	800	-	Publisher's	0
International Journal of General Medicine	2008	Born OA	Commercial	1865	-	Publisher's	50
Acta Crystallography: Structure Reports Online	2008	Converted to OA	Society	150	0.4	Publisher's	> 4000
Human Genomics and Proteomics	2009	Born OA	Commercial	1665	-	Publisher's	6

[a] The numbers of articles for 2010 have been determined by checking the journal websites. In the case of the larger journals, the numbers of articles are approximations.

XSL•FO
RenderX

Results

Journals Born as Open Access Founded by Individual Academics

The majority of journals that started publishing as open access during the mid- and late-1990s were new electronic-only journals (often with *electronic* or *online* as part of the name) founded by individual academics or groups of academics. The setting up of a new electronic-only OA journal was simple and required little infrastructure and capital particularly since there was no need for marketing to get subscribers. The central asset was the personal network of the editor, needed to recruit a credible editorial board and to solicit the first submissions. Usually the journals were hosted on the website of the editor-in-chief's university with home-crafted simple static Web pages. The management of the journal and the peer-review process were usually done on a voluntary basis, and the way the journals were operated was in spirit close to the way many open-source software development projects worked. The volume of manuscripts that were handled was usually rather low.

Medical Education Online is a good example of a pioneer volunteer-based OA journal [13]. The journal was from the start (1996) envisaged as a sort of portal for experts interested in medical education and also contained material other than peer-reviewed articles (ie, short discussion items, book reviews, and resource sections where academics could upload material), but over the years, the journal material has been more and more concentrated on articles. Accepted articles are published as they become ready rather than in regular issues, which speeds up publication. The look and feel of the articles is nevertheless exactly the same as in traditional scholarly paper journals.

Medical Education Online was originally launched with a number of invited articles, and for the first five years, the number of submissions and published articles was low. But after having survived the first critical years, the numbers have increased (currently around 20 published articles per year), and the journal has established itself within its research community.

For the first decade *Medical Education Online* was published using a Web platform programmed by the editor-in-chief. Over the years, the platform was improved to include, for instance, the possibility for electronic submission of manuscripts. Due to the increase in the workload, the journal adopted article-processing charges (APCs) in 2008 in order to generate a modest revenue. From the start of 2010, the journal has been published by a company specialized in open access publishing (Co-Action Publishing) and uses the Open Journal Systems (OJS) software, a widely used open-source solution for publishing scholarly journals and handling the review process [26]. The level of the APCs has been gradually raised to the current US $800 in order to cover the costs of professional copyediting and the costs of using a professional publishing firm.

The *Journal of Electronic Publishing*, founded in 1995, has had a slightly different development path. After some struggling first years, the journal was on hiatus for four years (2002-2005) before it reemerged and is now published by the University of Michigan Library [27]. Due to the sponsorship from the publishing organization, it has been able to avoid requiring article-processing charges.

Elore, is the oldest open access journal from Finland and is a good example of how scientific publishing in languages other than English can benefit from OA. It is published by the Finnish Folklore Society and operates with a minimal budget mainly using volunteer labor. Like many other similar journals, it has recently opted to take into use the OJS software. It publishes articles in both the national languages, Finnish and Swedish, but also publishes articles in English and includes items other than peer-reviewed articles.

Due to the strategic importance of maintaining the scientific discourse in national languages and promoting the local culture, governments and ministries in many countries are providing grants to support local scientific journals, particularly in the social sciences and humanities, where subscription journals are also struggling to make ends meet. In Finland, a problem for a long time was that these grants were based on a percentage of a journal's monetary income (usually from subscriptions), thus effectively excluding many OA journals from being eligible. Since 2006, the rules have been relaxed, and *Elore* has also benefitted from a small government grant.

In Canada, the Social Sciences and Humanities Research Council (SSHRC) has recently changed its rules for supporting scholarly journals so that the subsidy is Can $850 per published article, with a total maximum of Can $30,000 per journal and a ceiling of Can $5000 for paper or e-distribution costs. These regulations focus the support on the peer-review and copyediting costs of the journal production process, which means that OA journals get an equal treatment compared with paper journals. A very successful OA journal that has received a grant from SSHRC is the Toronto-based *Journal of Medical Internet Research* [15]. The *Journal of Medical Internet Research* is also a forerunner in experimenting with different sources of revenues, including submission fees in addition to charges for published articles, institutional memberships covering APCs of employees, fast-track handling of manuscripts for an extra fee, and sales of the PDF full text versions (the hypertext markup language [HTML] versions are OA) [15]. The journal also experiments with novel methods of peer-review (open peer review) and social media-based article level impact metrics (see Editorial in this issue).

Experimenting With Formats and Peer Review

In the mid-1990s, publishers of electronic journals assumed that most readers would prefer to read the articles on screen and would also prefer a straightforward HTML format for the articles, which, for instance, allowed direct hyperlinks to external Web references. Later on, many OA journals chose to format the articles as PDF files, which look like traditional articles in the printout format and which can be easily generated from word processing manuscripts. For the first decade, *Medical Education Online* published HTML and PDF versions in parallel, but since download statistics indicated that readers increasingly preferred PDF versions, the HTML format was dropped after 2005. *Information Research* [11], on the other hand, is still published in the HTML format although it has

XSL•FO
RenderX

recently adopted OJS for managing the review process and the article archive.

The *International Journal of Design Computing* (published between 1997 and 2003) dealt with a subject matter (architecture) where the possibility of including high class graphics, three-dimensional models, videos, and even virtual reality simulations in the material was expected to offer an important added value for readers. Like many other early OA journals, the *International Journal of Design Computing* dwindled after the first few years due to a lack of submissions. Academics seemed at that time reluctant to submit their best articles to new experimental electronic-only journals.

Diagnostic Pathology (begun in 2006) is another example of a journal trying to use the potential of the electronic medium. Authors can include virtual pathology slides with their articles, and readers can navigate in these with an easy-to-use viewing tool.

In the early years, there was also a lot of enthusiasm about trying out novel forms of peer review and commentary, which the Web enabled. The *Electronic Transactions on Artificial Intelligence* [28] experimented with a process in which authors first uploaded their manuscripts to the journal site followed by a period of open commentary by readers. After a few months, the author could request a formal anonymous peer review of the original submission or an improved version of the manuscript. If the article passed this peer review, which also took into consideration reader comments, it would then receive the status of published journal paper, but the results would have been disseminated earlier. The commentary from readers was also stored with the texts. This type of open peer review represents the linking of the functionality of a subject-based repository for preprints (such as arXiv for physics [29]) or the working paper tradition in disciplines like economics with a single peer-reviewed journal. Other early OA journals that have experimented with open peer review include the *Journal of Medical Internet Research* [15], discussed above.

In 2004, Copernicus Publications established the publication series *Hydrology and Earth Systems Science Discussions* as a complement to the existing journal, *Hydrology and Earth Systems Science* [30]**.** The idea is that discussion papers can be published within a few days in *Hydrology and Earth Systems Science Discussions* after only a very cursory screening by editorial board members. After that, reader comments and formal peer reviews are openly posted together with the discussed manuscripts. Those that pass the formal peer review are eventually published as full papers in *Hydrology and Earth Systems Science*. Currently, a dozen Copernicus journals use the same structure of twin journals and discussion forums. Copernicus is also interesting for its revenue model since it uses page charges for unrefereed manuscripts published in the discussion sections but publishes the ultimately accepted articles for free.

An additional way in which the Internet can be used to increase the transparency of the peer-review process is to upload the full prepublication history of the manuscript (the reviewers comments and the revisions of the manuscripts) together with the published article, a feature of *BMC Medicine*.

Postpublication peer review is currently being tried out by *Open Medicine*, which has started posting articles on wikis, open to changes and additions by readers [31]. The articles are of the review type and have first undergone a standard peer review before being posted. After that, readers can make changes and additions and also monitor changes and the document history. The idea is close to the idea behind Wikipedia articles, with the major difference being that the original seed document is of a peer-reviewed standard. Review articles are particularly suited to this type of treatment, since the state of the art is continuously changing as new research is being published.

Academics seem to be rather conservative in their choice of publication forums, particularly concerning peer-reviewed articles that are central elements in their publication lists. Due to this, the vast majority of open access journals still adhere to a rather conventional format, and peer-review practices remain largely unchanged.

Society Journals That Have Made the Electronic Version OA

A relatively low-risk route to OA has been for well-established printed subscription journals to make their electronic versions openly available. Very often the decision to do so has been taken at the time when the e-versions were first made available. One of the pioneers was *British Medical Journal***,** which started making its research articles openly available in 1998. *British Medical Journal* has a lot of advertising revenue, which is not affected by the decision, and it also offers other material, which is only open to subscribers.

The open e-version strategy has appealed in particular to society journals, which often are using electronic platforms from third parties. Strong society publishers have judged that they have a relatively stable subscription base and other income so that their subscription revenue would not suffer significantly, and they have at the same time been convinced of the service OA can offer the research community.

The leading third party e-portal for American and European society journals is Highwire Press. Among the 1527 journals currently using the platform, 282 offer delayed OA (usually by a year), including very high impact journals such as the *Journal of the American Medical Association*, *Brain*, and *European Heart Journal*. Another 48 of these 1527 journals offer immediate OA.

Outside the Anglo-American sphere, different types of e-portals have emerged. These portals are directly or indirectly government-sponsored and have a mission to help local scholarly journals reach a wider global audience. In a sense, they provide a form of subsidy for journals that choose to make their e-versions open access since their use is usually free provided that the journals fulfill scholarly criteria. Due to the economies of scale, these services are in fact quite cheap compared with the journals themselves setting up e-versions. Packer [32], for instance, mentions that the cost per published article is US $60 for the Scientific Electronic Library Online (SciELO) portal.

Such portals are very important in the Spanish and Portuguese speaking countries (SciELO, Red de Revistas Cientificas de

XSL•FO
RenderX

América Latina y el Caribe **.** [Redalyc]) [33] and in Japan (Japan Science and Technology Aggregator, Electronic [J-STAGE]) [34]. In total, these three portals alone contribute around 14% of all OA journals listed in DOAJ. A recent study has shown that of all the roughly 15,000 peer-reviewed journals indexed in 2010 in Scopus, the percentages that were OA were 73.9% for Latin America, 4.9% for North America, and 6.9% for Europe [35], clear evidence of how widely established high-quality Latin American journals had made their e-versions openly available via such portals.

As an example of the effects of such portals, consider the journal *Boletim do Museu Paraense Emílio Goeldi: Ciências Humanas*. This journal has its roots in one of the oldest scholarly journals in Brazil (established in 1894) and publishes articles in the social sciences and humanities with topics related to the Amazonas region. The language of the articles in the journal is Portuguese, but all articles also have abstracts in English. In fact, 15 years ago, finding out about articles in the journal would have been very difficult unless the reader belonged to a very select group of people who either had a personal subscription or their university happened to subscribe to and archive the paper journal. Since the electronic full text version of this article is now openly available via Scielo, anybody with Internet access who might take an interest in this sort of topic will now easily find it, for instance, via a Google keyword search or tracking a reference found in another publication.

Another example of the positive effects of third party OA portals on bridging the digital divide is University of Toronto-based Bioline International, which explicitly aims at helping journals in developing countries publish electronic OA versions. Bioline finances its operations via sponsorship and supporting members. One of the 54 journals on Bioline's website is the *African Journal of Food, Agriculture, Nutrition and Development* published by the Rural Outreach program based in Kenya. The journal also has a print edition, but has been available since 2005 in an electronic OA version via Bioline. Considering the global challenges in feeding the world's inhabitants, the journals' articles can be of interest to a wide audience in academia, government, and international organizations.

The Emergence of Specialized OA Publishers

BioMed Central (BMC), founded in the year 2000 by Vitek Tracz, was the first specialized professional OA publisher. Since 2002, the business model of BMC has been to fund operations mainly with article processing charges (APCs) and to launch a large number of journals in different fields of biology and medicine to benefit from economies of scale in e-infrastructure, marketing, and so on. BMC was eventually successful enough to attract the large mainstream publisher Springer to buy the company in 2008. In 2010, BMC's 234 journals published more than 15,000 articles.

An interesting example of a BMC journal is the *Malaria Journal***,** which publishes research on topics of vital interest to researchers and practitioners in the developing world, who often have problems financing subscriptions to the research literature. Hence, open access is of particular importance. BMC, as most OA publishers, waives the APCs (currently US $1775 for *Malaria Journal*) for authors who have problems getting funding for this, in particular authors from developing countries.

Another BMC journal with an innovative scope is the *Journal of Negative Results in Biomedicine* [36]. Over the nine years of its existence, the journal has published few articles and thus cannot be considered successful, but probably due to its low marginal costs and APC revenue, it is still operating.

Public Library of Science was originally mainly a Web campaign promoting open access. When the campaign failed to have the intended impact, the originators together with Harold Varmus, a Nobel Prize winner and former director of the National Institutes of Health, founded an OA publishing company, also named Public Library of Science (PloS). Thanks to a substantial initial grant of US $9 million, the company was able to launch two very high quality journals in 2003-2004 and has since expanded to seven journals. *PLoS Biology* currently has the highest ISI impact factor (12.9) of all general biology journals. In addition to the fast peer-review and publishing schedules typical for OA journals, PLoS has strived to offer both authors and readers articles with high-class layout and interactive features, including download statistics and reader comments.

Both BioMed Central and PloS publish journals mainly in biomedicine, a segment of science where research funding is abundant and where authors (through their institutions) can usually afford to pay the APCs. Other OA publishers try to cover all fields of science with their journal portfolios. The publisher Hindawi is an interesting case, since it operates from Egypt and has been able to keep publishing costs down due to much lower personnel costs [14,37]. Despite this, its operations are fully global. Hindawi was founded as a conventional publisher in 1997 but started to convert journals to OA financed with APCs in 2003, and four years later, all of its journals were OA. A good example of Hindawi's journals is *Advances in Difference Equations*. For a journal in mathematics, the peer review and copyediting can be quite labor-intensive, but the APC is still quite reasonable at US $600. The journal is a popular outlet for mathematicians from a wide spectrum of countries, and its global reach is well reflected in the composition of its scientific editorial staff and editorial board.

The picture of open access publishers wouldn't be complete without a discussion of Bentham and Dove Press, both of which have created controversy in the OA publishing debate [38,39]. Bentham massively launched over 200 OA journals in 2007 under the label Bentham Open. In connection with the launch, academics around the world were spammed with emails offering membership in editorial boards and soliciting submissions.

In 2009, Phil Davis reported that he and a colleague had submitted a grammatically correct but nonsensical manuscript generated by a software program to Bentham's *Open Information Science Journal* and that he had subsequently received a mail stating that the article had been accepted for publishing provided he would first pay the publication charge of US $800. After some media coverage of the scandal, the editor-in-chief of the journal resigned, claiming that he had no knowledge of the manuscript in question and its acceptance [40].

XSL•FO
RenderX

An example of Dove Press journals is the *International Journal of General Medicine*. Articles in this journal have a very professional visual look. Highly visible statistics on the journal home page promise that the average time from submission to acceptance (including peer review) is 13 days with an additional wait of 15 days until final publication (as of the February 12, 2011).

Bentham and Dove Press seem to have identified a niche market of academic authors who are in desperate need of rapidly getting manuscripts of possibly questionable scientific quality published in journals, which can be labeled peer-reviewed and who are willing to pay the required article processing charges. It is still not clear if either of these publishers will succeed in making this a profitable and sustainable operation.

Converting Journals From the Subscription Model to Open Access

The vast majority of OA journals are either newly created electronic-only journals or established journals, which make their electronic versions available but finance their operations with income from their printed versions. Converting a subscription journal to full OA is much riskier, particularly if the journal will be funded with APCs, and for this reason, there are few such cases. One example of a successful transition is the Oxford University Press (OUP) journal *Nucleic Acids Research*. The conversion in 2004 was part of broader OA strategy in which a number of OUP journals started allowing authors to open up individual articles in subscription journals against payment [41]. *Nucleic Acids Research* was chosen for the conversion because it was already well established as a quality journal (currently in the top 10% of its field with an ISI impact factor of 7.4); hence, the risk of submissions dwindling away after the conversion to APC funding was deemed to be low.

A totally different type of conversion might come about through pressure from major subscribers. A number of the biggest nuclear research institutes in the world, including the European Organization for Nuclear Research (CERN), have founded the Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP3), which aims to force the major physics journals to switch to the open access model. The consortium is currently collecting pledges from potential additional consortium members and by December 20011 had collected 80% of the €10 million they estimate will be required to buy the OA publishing services of the journals in question. In particle physics, a few huge laboratories contribute a major part of the subscription income of the leading journals in the field. The participants in the consortium would offset their contributions by canceling their subscriptions to the targeted journals.

Experimenting With Gradual Introduction of OA by Introducing Hybrid Journals

Due to the high commercial risks in converting established journals to OA, several major publishers have introduced so-called hybrid journals, traditional subscription journals that allow authors—for a payment—to make their individual articles openly accessible. From the publisher's viewpoint, this has been a risk-free experiment with OA. Springer pioneered this in 2004 with an open access program known as Open Choice, and other publishers that have followed include Oxford University Press with an open access program know as Oxford Open and Sage with an open access program known as Sage Open. Springer's initial choice of US $3000 as a uniform price for Open Choice across all journals seems to have set a price standard followed by others. The low uptake of the hybrid model, 1% to 2 % of eligible articles [24], indicates both that the level of the charges might have been too high compared to the benefits the authors perceive they get and also that a uniform pricing model across a large portfolio of journals doesn't work.

A recent development in the last couple of years is that major mainstream publishers have also started launching new full OA journals. In addition to the purchase of BioMed Central, Springer has, for instance, recently launched 32 full open access journals under the label Springer Open.

Mega Journals

A new type of journal that has emerged recently is the *Mega* journal, publishing several thousand articles per year over a broad spectrum of topics. The primary example of this type is *PLoS ONE*, which accepts manuscripts in any field of science or medicine. In addition to the broad scope, *PLoS ONE* introduced an important change to the function of the peer review. This change is best explained by a direct quote from the journal web site [42]:

> *Too often a journal's decision to publish a paper is dominated by what the editor/s think is interesting and will gain greater readership—both of which are subjective judgments and lead to decisions which are frustrating and delay the publication of your work. PLoS ONE will rigorously peer review your submissions and publish all papers that are judged to be technically sound. Judgments about the importance of any particular paper are then made after publication by the readership (who are the most qualified to determine what is of interest to them)*

To support the idea that it is the readers that eventually will determine the importance and the *contribution* of any particular article, *PLoS ONE* utilizes interactive tools for readers and metrics such as downloads and citations per paper. The download statics show a highly skewed distribution between a vast majority of articles that have less than average readership and a small minority of articles that are widely read and cited. *PLoS ONE* thus seems to be succeeding in combining the dissemination function of a subject-based preprint repository such as arXiv and the quality certification function of traditional journals. Since 2010, *PLoS ONE* has had an impact factor of 4.3. In only five years, it has rapidly increased its publication volume to over 10,000 articles per annum.

Recently, several mainstream science publishers have launched this type of journal, for instance, *Sage Open* for the social sciences and humanities, *Nature Scientific Reports* for the natural sciences, *BMJ Open* for medicine and the Royal Society´s *Open Biology*.

Acta Crystallography: Structure Reports Online represents a totally different type of mega journal. Published by the

XSL•FO
RenderX

International Union of Crystallography, it was originally, like a number of other journals, a subscription journal but converted to OA in 2008 [43]. The APC is low at only US $150 as compared with US $1300 at *PLoS ONE*. *Acta Crystallography: Structure Reports Online* publishes short, highly structured articles in an extremely narrowly field. As the publication volume has increased (over 5000 articles in 2008) the journal archive has begun to look more and more like a database of scientific data.

Acta Crystallography: Structure Reports Online also highlights one aspect of publishing where the difference between subscription and OA journals is accentuated. In subscription journals, the rights of authors and readers are highly restricted in order to protect the commercial interests of the publishers. In OA journals, there is no such interest at stake, and publishers mostly allow the authors to retain the copyright. Furthermore, OA publishers have increasingly started to adopt creative commons (CC) licenses, which quite explicitly regulate the rights of readers (including software tools) to use and reuse the publications and which include computer readable versions as a standard feature. The CC licenses are related to the licenses (ie, GNU) used for open-source software.

The CC licenses are vital in promoting the reuse of published research data, in particular in the sciences. *Acta Crystallography: Structure Reports Online*, for instance, allows authors to include structured data sets in the refereed articles, using a standardized syntax called crystallographic information file (CIF). Murray-Rust [44] has demonstrated how such data from uncoordinated articles found on the Web can subsequently be harvested by data mining tools to form a knowledge base of much greater power than the isolated articles.

Other publishers are also experimenting with linking journal articles and datasets. *Human Genomics and Proteomics*, which started publishing in 2009, is a joint venture of Sage and Hindawi and encourages authors to publish data sets that will be stored in an open repository called FINDbase, a population-specific genetic database that charts causative mutation frequencies and their associated disorders in several countries around the world [45]. An author can submit the dataset and an abstract about it for peer review in *Human Genomics and Proteomics* [46]. After acceptance, the abstract is published in the journal. The Genetics Society of America has also recently announced a new OA journal called *G3: Genes|Genomes|Genetics*, which particularly aims to encourage the inclusion of large structured data sets in the articles.

Discussion

Different Types of Innovative Features

The journals discussed in this article provide evidence of the opportunities for innovation that open access provides. The rest of this discussion focuses on the different business models used to achieve sustainability for OA publishing and on additional features made possible or facilitated by OA and the electronic format. Table 3 below shows the innovative features discussed earlier in this article in the Introduction as well as which journals are particularly good examples of the use of the feature in question. This does not mean that the journals listed don't also have the feature in question, for instance, most of the born-OA journals have faster publication schedules than traditional journals. The central innovation has, of course, been the open access as such, and most OA journals have primarily focused on achieving this, with few changes in article formats, peer-review practices, and so on.

Table 3. Innovative features discussed in the article and journals that provide good examples of each feature.

Innovative Feature	Example Journals
Universal accessibility	All
Cost-savings by using volunteers	*Elore, Information Research*
Cost savings by using open-source software	*Medical Education Online, Information Research*
Cost savings by using third party e-portals	*Boletim do Museu Paraense Emílio Goeldi*
Funding by article processing charges	Journals from Plos, BMC, Bentham, and DovePress, and the following: *Journal of Medical Internet Research, Medical Education Online, Nucleic Acids Research*
Broader or narrower journal topics due to the global reach	*PLoS ONE, Journal of Negative Results in Biomedicine*
Novel peer-review methods	*Electronic Transactions on Artificial Intelligence, Hydrology and Earth System Sciences, PLoS ONE, Journal of Medical Internet Research*
Faster article publication cycles	*International Journal of General Medicine, PLoS ONE, Journal of Medical Internet Research*
More flexibility in the layout and structure of articles	*The International Journal of Design Computing, Diagnostic Pathology*
Interactivity for after-publication discussions	Journals from PLoS and BMC
Easy reusability of the digital content	*Acta Crystallography: Structure Reports Online, Human Genomics and Proteomics*

XSL•FO
RenderX

New Ways of Saving Costs and Getting Revenue

Many well-established print journals have been able to rely on their stable subscription income to open up their electronic versions for no charge. This option has been quite common among society journals, in particular among journals from non-English-speaking countries, which in many cases have been able to use national e-portals for free.

There are a number of ways of financing electronic-only OA journals. Many of the early pioneering journals relied on voluntary labor and use of the website of the editor's university free of charge. This has worked quite well for journals handling small numbers of submissions and publishing on technically simple websites but has not worked well as the number of articles has increased. There are examples of successful early journals that have later adopted article-processing charges in order to ensure continuity and a more professional way of operating. Many independent and society journals have adopted Open Journal Systems software as a cost-effective way of getting a fairly robust IT platform that also incorporates manuscript-handling features.

As professional OA publishers have entered the field, the article processing charge has become the central funding mechanism for large-scale full OA publishing. More and more, it is the quality level of the journal that determines the article processing charge that authors (and their funders) are willing to pay, especially when considering alternatives inside one's research discipline. In addition to quality, the subject field also affects the possibilities and willingness of authors to pay. APCs in biomedicine are typically higher than in the social sciences, for example.

OA journals that have adopted article-processing charges have almost exclusively levied these for published articles. This means that authors whose manuscripts have been accepted have indirectly paid the costs incurred for rejected manuscripts. This choice has obviously been based on the assumption that charging even a small amount for submissions might stem the inflow of manuscripts.

Normally APCs are the same for all articles within a given journal, but a nice feature of the major OA publishers is that they usually promise to waive the charges for authors who can document that they have problems financing APCs, especially authors from developing countries.

Broader or Narrower Journal Topics Due to Global Reach

Open access seems to be particularly well suited to what could be called "microtopics" and "macrotopics." Open access has clearly lowered the threshold for founding new journals in narrow areas, which in the print and subscription model would not have been economically viable. Understandably, many journals founded in the 1990s specialized in topics related to IT and the Internet.

Likewise, open access offers an excellent way for journals from countries outside the major Anglo-Saxon sphere, both those publishing in English and those publishing in other national languages, to increase their readership and impact. Hence, OA lowers the digital divide by allowing scientists in developing nations both better access to mainstream science and increased chances of being read outside their own countries.

A recent trend is the emergence of mega journals, the topics of which span substantial parts of all of science. Of the journals considered, mega journals were published by well-established, credible publishers with professional staff and ready IT infrastructure. The key issue for such journals in particular is the ability to attract submissions, manage reviews, and recruit reviewers.

Novel Peer-Review Methods

Several OA journals have experimented with different variations of open peer review, which relies on the activity of readers to actively upload comments to the journal websites and which allows the research results to be made public at the preprint manuscript stage. So far, the results are inconclusive, and open peer review is still quite rare.

There is an obvious temptation for some commercial OA publishers so set up journal collections that can publish submissions with minimal costs and efforts for the peer review. While this is fully legal as a business model, the scientific community can ignore giving much credit to such publications in its evaluations. Early evidence also suggests that such publishers have had problems getting enough submissions.

The *review of scientific rigor only*, a concept that *PLoS ONE* has pioneered, and where the scientific contribution is determined by readership and citations rather than the judgment of a couple of peer reviewers, seems on the other hand a very useful innovation, at least as a model for mega journals with broad scope.

Faster Article Publication Cycles

Faster publication has always been an advantage of open access journals, in particular for journals that are not published in an issue format. Some journals have recently streamlined their processes in order to achieve very short average lead times from submission to publication (of accepted papers). The very short average processing times announced by Dove Press, however, raise many questions concerning the quality of the review process.

The electronic-only format freed OA journals from the straightjacket of the journal issue, and many OA journals have from the start opted for publishing articles on the fly as the articles become technically ready. This was seen early on as one of the major benefits of OA since it speeded up publication, usually by several months. Lately, traditional journals have partly followed this lead by making *articles-in-press* available to subscribers on their websites.

More Flexibility in the Layout and Structure of Articles

The electronic format has also opened up new possibilities for including types of presentation formats other than the linear text format, particularly in OA journals, which don't have the burden of also being published in print. Media enhancements as well as documentation attached to the articles have also been tried, but such additions may present problems for peer

XSL•FO
RenderX

reviewing and can be tricky to handle for journal archives. On the whole, authors seem, however, to be rather conservative concerning changes from the traditional look and feel of articles.

Of the different types of material that could not easily fit into traditional printed articles, the structured data set is the most promising, particularly in some domains of science where such data can be data mined and harvested into bigger aggregate services.

Interactivity for After-Publication Discussions

The electronic format offers opportunities for new kinds of interactive functionality, which was not possible in printed journals. Since OA journals were the first electronic scholarly journals, it was natural that they first started to experiment with reader comments, open peer review, and so on. Other interesting features that are included nowadays in PLoS journals, for instance, are download and citation metrics. Subscription journals can also include such features now, but only in the electronic versions.

Easy Reusability of the Digital Content

For open access journals, the assignment of copyright and the licensing agreements for readers and automated tools differ radically from traditional subscription-based journals. During the 1990s, the OA journals were mostly just *open* and the copyright and license terms were usually not formalized. The Creative Commons standard licenses for Web material emerged after the year 2000 and are eminently suited for scientific publications and data. Currently, most professional OA publishers use some form of CC license, and its use is also spreading among independent OA journals. The increasing use of Creative Commons licenses in OA journals facilitates, in particular, the data mining of data attached to articles.

Conclusions

Open access publishing is rapidly increasing its share of the overall volume of scientific journal publishing with an annual growth rate of 20% and an estimated number of more than 250,000 articles in 2011 (extrapolated from [10]). So far, this growth has almost exclusively come from independent, society, and newly started OA publishers. Now the tide seems to be turning. The fact that major scholarly publishing companies are in the process of launching new APC funded journals is a clear indication that they have judged that the OA model has proved to be sustainable. Existing OA journals have already tried out many new ideas in scholarly publishing, as reported in this paper. The successful innovations are fast becoming part of the academic infrastructure, with scientists voting with their manuscripts as to which ones will prevail.

Conflicts of Interest

None declared

References

1. Ware M, Mabe M. The STM report: An overview of scientific and scholarly journal publishing. Oxford, UK: International Association of Scientific, Technical and Medical Publishers; 2009. URL: http://www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf [WebCite Cache ID 61ythDH8T]
2. Kircz J, Roosendaal H. Understanding and Shaping Scientific Information Transfer. : UNESCO; 1996 Presented at: Joint ICSU Press/UNESCO Expert Conference on Electronic Publishing in Science; 19-23.2.1996; Paris URL: http://eos.wdcb.ru/eps1/kircz.htm
3. Gierveld H. A Conceptual Analysis of Functions, Processes and Products in the Scholarly Communication Chain. 2002 Presented at: Proceedings of the 6th International ICCC/IFIP Conference on Electronic Publishing; 6-8.11.2002; Karlovy Vary, Czech Republic URL: http://elpub.scix.net/cgi-bin/works/Show?02-16
4. Benos DJ, Bashari E, Chaves JM, Gaggar A, Kapoor N, LaFrance M, et al. The ups and downs of peer review. Adv Physiol Educ 2007 Jun;31(2):145-152 [FREE Full text] [doi: 10.1152/advan.00104.2006] [Medline: 17562902]
5. Dewatripont M, Ginsburgh V, Legros P, Walckiers A, Devroey JP, Dujardin M, et al. Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe. Brussels, Belgium: Directorate-General for Research, European Commission; 2006. URL: http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf [accessed 2011-09-26] [WebCite Cache ID 61yubAc1J]
6. Harnad S, Brody T, Vallières F, Carr L, Hitchcock S, Gingras Y, Hilf E. The Access/Impact Problem and the Green and Gold Roads to Open Access. Serials Review 2004;30(4):310-314.
7. Houghton J, Rasmussen B, Sheehan P, Oppenheim C, Morris A, Creaser C, et al. Economic implications of alternative scholarly publishing models: Exploring the costs and benefits. London, UK: Joint Information Systems Committee, JISC; 2009. URL: http://ie-repository.jisc.ac.uk/278/3/EI-ASPM_Report.pdf [accessed 2011-11-28] [WebCite Cache ID 63XJsKH8K]
8. Matsubayashi M, Kurata K, Sakai Y, Morioka T, Kato S, Mine S, et al. Status of open access in the biomedical field in 2005. J Med Libr Assoc 2009 Jan;97(1):4-11. [doi: 10.3163/1536-5050.97.1.002] [Medline: 19159007]
9. Björk BC, Welling P, Laakso M, Majlender P, Hedlund T, Gudnason G. Open access to the scientific journal literature: situation 2009. PLoS One 2010;5(6):e11273 [FREE Full text] [doi: 10.1371/journal.pone.0011273] [Medline: 20585653]
10. Laakso M, Welling P, Bukvova H, Nyman L, Björk BC, Hedlund T. The development of open access journal publishing from 1993 to 2009. PLoS One 2011;6(6):e20961 [FREE Full text] [doi: 10.1371/journal.pone.0020961] [Medline: 21695139]

XSL•FO
RenderX

11. Wilson T. Trials and Tribulations? Editing Information Research, an Open Access Electronic Journal. Sciecom Info 2005;2(3) [FREE Full text]
12. Björk BC, Turk Ž. The Electronic Journal of Information Technology in Construction (ITcon): an open access journal using an un-paid, volunteer-based organization. Information Research 2006;11(3):255 [FREE Full text]
13. Solomon D. Medical Education Online: a case study of an open access journal in health professional education. Information Research 2007;12(2):301 [FREE Full text]
14. Maron N, Smith K, Loy M. Hindawi Publishing Corporation: The Open Access Contributor Pays Model. In Sustaining Digital Resources: An On-the-Ground View of Projects Today. Document No 609:73-80. London, UK: Joint Information Systems Committee, JISC; 2009. URL: http://www.ithaka.org/ithaka-s-r/research/ithaka-case-studies-in-sustainability/report/SCA_Ithaka_SustainingDigitalResources_Report.pdf [WebCite Cache ID 61yuoSVed]
15. Eysenbach G. 10 years experience with pioneering open access publishing in health informatics: the Journal of Medical Internet Research (JMIR). Stud Health Technol Inform 2010;160(Pt 2):1329-1333. [Medline: 20841900]
16. Tidd J, Bessant J, Pavitt K. Managing innovation: integrating technological, market and organizational change. Chichester: John Wiley & Sons; 2001.
17. Tenopir C, King D. Towards electronic journals: realities for scientists, librarians, and publishers. Washington, DC: Special Libraries Association; 2000.
18. Björk BC. A model of scientific communication as a global distributed information system. Information Research 2007;12(2):307 [FREE Full text]
19. Baregheh A, Rowley J, Sambrook S. Towards a multidisciplinary definition of innovation. Management decision 2009;47(8):1323-1339.
20. Teece, D. Business Models, Business Strategy and Innovation. Long Range Planning 2010;43(2-3):172-194.
21. Ziobrowski A, Gibler K. Factors Academic Real Estate Authors Consider When Choosing Where to Submit a Manuscript for Publication. Journal of Real Estate Practice and Education 2000;3(1):43-54.
22. Swan A, Brown S. Authors and open access publishing. Learned Publishing 2004;17(3):219-224.
23. Björk BC, Öörni A. A Method for Comparing Scholarly Journals as Service Providers to Authors. Serials Review 2009;35(2):62-69.
24. Dallmeier-Tiessen S, Goerner B, Darby R, Hyppoelae J, Igo-Kemenes P, Kahn D, et al. Open Access Publishing - Models and Attributes. Munich, Germany: Max Planck Digital Library; 2010. URL: http://edoc.mpg.de/478647 [accessed 2011-09-26] [WebCite Cache ID 61yuvag4S]
25. Directory of Open Access Journals. URL: http://www.doaj.org/ [accessed 2011-11-08] [WebCite Cache ID 63ZzyR3Na]
26. Edgar B, Willinsky J. A survey of scholarly journals using Open Journal Systems. Scholarly and Research Communication 2010;1(2):1-22 [FREE Full text]
27. Bonn M. Why we publish JEP. Journal of Electronic Publishing 2006;9(1). [doi: 10.3998/3336451.0009.102]
28. Sandewall E. Publishing and reviewing in the ETAI. Electronic Transactions on Artificial Intelligence 1997;1:1-12 [FREE Full text]
29. arXiv.: Cornell University Libary URL: http://arxiv.org/ [accessed 2011-11-08] [WebCite Cache ID 63a0KxmD2]
30. Roth, K, Savenije J, Sivapalan, M. Editorial. Hydrology and Earth System Sciences Discussions 2004;1:1-5 [FREE Full text]
31. Murray S, Giustini D, Loubani T, Choi S, Palepu A. Medical research and social media: Can wikis be used as a publishing platform in medicine? Open Med 2009;3(3):e121-e122. [Medline: 21603044]
32. Packer A. The SciELO Open Access: a gold way from the south. Canadian Journal of Higher Education 2009;39(3):111-126 [FREE Full text]
33. Redalyc. Red de Revistas Científicas de América Latina y el Caribe, España y Portugal URL: http://redalyc.uaemex.mx/ [accessed 2011-11-08] [WebCite Cache ID 63a0RCUEC]
34. J-STAGE. Japan Science and Technology Aggregator, Electronic URL: http://www.jstage.jst.go.jp/browse/ [accessed 2011-11-08] [WebCite Cache ID 63a0TUoDg]
35. Miguel S, Chinchilla-Rodríguez Z, Moya-Anegón F. Open access and Scopus: a new approach to scientific visibility from the standpoint of access. Journal of the American Society for Information Science and Technology 2011;62(6):1130-1145.
36. Pfeffer C, Olsen BR. Editorial: Journal of negative results in biomedicine. J Negat Results Biomed 2002;1:2 [FREE Full text] [Medline: 12459050]
37. Peters, P. Going all the way: how Hindawi became an open access publisher. Learned Publishing 2007;20(3):191-195. [doi: 10.1087/095315107X204049]
38. Davis P. Open Access Publisher Accepts Nonsense Manuscript for Dollars, The Scholarly Kitchen, 10th June. 2009. URL: http://scholarlykitchen.sspnet.org/2009/06/10/nonsense-for-dollars/ [accessed 2011-09-26] [WebCite Cache ID 61yvN2ZS1]
39. Greenberg C. Amy Bishop, Dove Press, and a publisher's conversion to Open Access. Blog posting, Open Biomed Info. 2010 Mar 01. URL: http://openbiomed.info/2010/03/amy-bishop-dove-press-and-a-publishers-conversion-to-open-access/ [accessed 2011-09-26] [WebCite Cache ID 61yvTRB6O]

XSL•FO
RenderX

40. Gilbert N. Editor will quit over hoax paper: Computer-generated manuscript accepted for publication in open-access journal. Nature News. 2009 Jun 15. URL: http://www.nature.com/news/2009/090615/full/news.2009.571.html [accessed 2011-11-08] [WebCite Cache ID 63a0aSHPm]
41. Bird, C. Oxford Journals' adventures in open access. Learned Publishing 2008;21(3):200-208.
42. PLoS ONE. PLoS ONE Journal Information URL: http://www.plosone.org/static/information.action [accessed 2011-11-08] [WebCite Cache ID 63a0ceVeS]
43. Baker G. Acta Crystallographica E converts to OA. Open Access News. 2008 Jul 03. URL: http://www.earlham.edu/~peters/fos/2008/03/acta-crystallographica-e-converts-to-oa.html [accessed 2011-09-26] [WebCite Cache ID 61yvclWMx]
44. Murray-Rust, P. Open Data in Science. Serials Review 2008;34(1):52-64.
45. Harris, S. Journal unites research articles with raw data, Research Information Aug/Sept. 2008. URL: http://www.researchinformation.info/features/feature.php?feature_id=186 [accessed 2011-11-08] [WebCite Cache ID 63a0fkdVM]
46. Patrinos GP, Petricoin EF. A new scientific journal linked to a genetic database: towards a novel publication modality. Hum Genomics Proteomics 2009;1. [doi: 10.4061/2009/597478] [Medline: 20948565]

Abbreviations

APC: article processing charge
BMC: BioMed Central
CERN: European Organization for Nuclear Research
SCOAP3: Sponsoring Consortium for Open Access Publishing in Particle Physics
CC: creative commons
CIF: crystallographic information file
HTML: hypertext markup language
ISI: Institute for Scientific Information
IT: information technology
J-STAGE: Japan Science and Technology Aggregator, Electronic
OA: open access
OJS: Open Journal Systems
OUP: Oxford University Press
Redalyc: Red de Revistas Cientificas de América Latina y el Caribe
SciELO: Scientific Electronic Library Online
SSHRC: Social Sciences and Humanities Research Council

Edited by G Eysenbach; submitted 12.03.11; peer-reviewed by J Willinsky, C Sutton; comments to author 10.05.11; revised version received 06.08.11; accepted 25.09.11; published 16.12.11.

Please cite as:
Björk BC
A Study of Innovative Features in Scholarly Open Access Journals
J Med Internet Res 2011;13(4):e115
URL: http://www.jmir.org/2011/4/e115/
doi:10.2196/jmir.1802
PMID:22173122

©Bo-Christer Björk. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 16.12.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

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

XSL•FO
RenderX