Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 16 (2014), Issue 2   ISSN 1438-8871   Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Editorial

Viewpoints

Original Papers

Reviews

Original Paper

Editorial

Harnessing the Social Web for Health and Wellness: Issues for Research and Knowledge Translation

Kendall Ho[1], MD, FRCPC; Peter Wall Workshop Participants[2]

[1]Faculty of Medicine, eHealth Strategy Office, University of British Columbia, Vancouver, BC, Canada
[2]see acknowledgements, Vancouver, BC, Canada

Corresponding Author:
Kendall Ho, MD, FRCPC
Faculty of Medicine
eHealth Strategy Office
University of British Columbia
Room 215
855 West 10th Avenue
Vancouver, BC, V5Z 1L7
Canada
Phone: 1 6048754111 ext 69153
Fax: 1 6048755083
Email: Kendall.Ho@ubc.ca

Abstract

Social media is a powerful, rapid, and popular way of communication amongst people around the world. How can health professionals and patients use this strategy to achieve optimal disease management and prevention and attainment of wellness? An interdisciplinary group at University of British Columbia, supported by a grant from UBC Peter Wall Institute of Advanced Studies, conducted a research workshop in February 2013 to explore what is known and yet to be researched in using social media for nurturing the growth of virtual communities of people for health and wellness. This two and a half day workshop brought together a group of 30 multidisciplinary experts in closed discussions to reflect on five research themes in detail: (1) individual information acquisition and application, (2) community genesis and sustainability, (3) technological design issues, (4) knowledge management, dissemination, and renewal, and (5) research designs. In addition, a public forum for the general public, which attracted over 195 live participants, over 100 participants via Web casting, 1004 tweets, and 1,124,886 impressions following the #HCSMForum hash tag on Twitter, demonstrated the keen interest of the general public in this topic. Key concepts were captured in JMIR publications in this issue, and future directions, including research, knowledge translation approaches, and strategic partnerships of interdisciplinary researchers with policy makers and industries emerged from the workshop proceedings.

(J Med Internet Res 2014;16(2):e34) doi:10.2196/jmir.2969

KEYWORDS

social media; research; interdisciplinary

Introduction

The Journal of Medical Internet Research (JMIR) has recently published a small theme issue with three papers that emerged from a workshop on social media [1-3]. Social media is a collective term that describes a rapidly developing group of powerful and ubiquitous technologies and a set of socio-technical approaches for people to connect, support, and learn from each other [4]. Social media provides a powerful means of connectivity that can join community members in active pursuit of issues of common interest [5]. Dialogues of virtual communities formed through social media can bypass traditional methods of peer review, knowledge synthesis, and validation, thereby accelerating the dissemination and democratization of information by novices and experts alike [6-8]. Concerns have been raised that the misuse of social media forums can lead to spread of inaccurate or even harmful information, and potentially misguide participants into doing more harm than good to their wellness [9,10].

We need to understand how this popular and emerging phenomenon of social media can be effectively and safely leveraged for the pursuit of wellness and health of a community and its members. This popular and emerging phenomenon of health management calls for illumination [11]. Research is needed on how the use of social media for health and wellbeing differs across different contexts (eg, different social norms, individual attitudes and practices, professional interventions, reward structures, and learning and play) [12]. Also important is examining how health information dissemination and health practices unfold and evolves fluidly with the to and fro of interaction among participants in social media communities, contrasting this with systems designed for planned application. We need to examine the operation of new designs in social media to explore and gain a greater understanding of the impact of online tools and concepts such as open source, peer production, crowdsourcing, virtual communities, participatory culture, e-learning, and game play, on the wellbeing and health of their users.

A multi-disciplinary approach would benefit this examination, studying the issue through different research traditions, including but not limited to the social sciences (sociology, anthropology, psychology, communications), knowledge science (library science, education), technologies (engineering, computer science), management science (business engineering), and health sciences (medicine, nursing, public health). Collectively, studies from different research areas, such as social informatics, the science of socio-technical systems, eHealth, e-science, e-learning, management information systems and human-computer interaction, information science, computer science, management science, sociology, psychology, and anthropology will provide a more complete picture on the topic.

The aim of our workshop, “Harnessing the Social Web: Communities for Health and Wellness” and the study was to gather an interdisciplinary group of researchers to examine the issues of using social media for supporting health and wellness.

Scope of the Peter Wall Workshop

A group from 5 faculties at the University of British Columbia (UBC)—Medicine, Library Science, Nursing, Sauder School of Business, and Engineering—successfully obtained a competitive grant funding from the UBC Peter Wall Institute of Advanced Studies to organize a research workshop to invite experts in various disciplines who share a common interest in social media and the use of the Web for wellness and health to explore the various dynamics at play in social media that contribute to success or failure, and the related management of information, social support, decision-making, and action in support of individual and community wellness goals. The purpose of the workshop was to bring together a group of experts to pool knowledge in support of a multi-faceted understanding of social media use in health and wellness by and for individuals and communities to explore 6 research themes (see Textbox 1).

Textbox 1. The 6 research themes discussed at the Peter Wall workshop.

1. *Individual information acquisition and application:* How do individuals obtain, disseminate, and trust information obtained via social media, and how do these interactions influence change in attitude, commitment, knowledge, and behavior over time? How does the public nature of this information affect willingness to contribute? How do different cultural approaches to health and wellbeing influence individual and group behaviors?
2. *Community genesis and sustainability:* What are the dynamics that lead to the successful formation, growth, and sustainability of the virtual community facilitated by social media, and what different types of roles do individuals play to animate and enliven this community (leaders, lurkers, contributors, knowledge experts, etc)?
3. *Technological design issues:* What are the design strategies used in mobile phones, tablets, or the Web that create an optimal and positive experience for participants of the social media? For example: ease of access, quality of communication, speed of connectivity and optimal decision support to encourage positive change and ethical attention to values, privacy, etc. What are the ethical considerations in design? How should such technologies be designed to preserve the privacy, anonymity, confidentiality of individuals and their contributed data?
4. *Stakeholders' mutual influence towards wellness:* How do different stakeholders interact amongst and between themselves (eg. health professionals, health consumers) towards change, and what effective social phenomenon is at play in facilitating interaction among different groups? How has consumer knowledge shifted the authority previously held by health care providers?
5. *Knowledge management, dissemination, and renewal:* How would the information in the social media dialogues merge and contribute to evidence based information in support of local and outcome-specific contexts? How would this information be best represented and visualized such that virtual community members can understand and apply it? How can we make best use of two-way flow of information and experience from novices to experts, and from individuals to crowds and communities?
6. *Research designs*: What methodological approaches and research traditions provide the right kind of design and ethical practices for examining social media use by an ecology of patient and health care providers, and for capturing the data necessary to address the questions above?

Workshop Program Overview

We successfully invited a group of 30 multidisciplinary experts from library and information science, medicine, engineering, computer science, nursing, public health, communications, anthropology, education and sociology, health policy, and business and health economics to participate in this workshop, which took place from February 3 to 5, 2013 (Figure 1). The experts were engaged to consider the 6 research themes (Textbox 1), and produced background papers in preparation for the workshop. We also engaged graduate students from interdisciplinary studies, including the School of Population and Public Health, School of Nursing, School of Library, Archival and Information Studies, and Department of Computer Science to participate in our discussion and contributed their ideas and insights to our discussion.

During the three days of the workshop, we maximized interactions amongst participants in order to draw upon their own disciplines' perspectives and experiences on social media for health and identify interdisciplinary opportunities and gaps needing further illumination and research. We concentrated on generating academic papers to capture the multi-dimensional ideas discussed at the workshop, resulting in 3 full papers and this editorial in the Peter Wall workshop theme issue in JMIR [1-3].

On the evening of February 4th, 2013**,** we conducted a public forum to present issues of using social media for wellness, inviting a few members from our workshop to give a keynote presentation and to be panellists of the forum. We invited the audience to engage in a dialogue to explore issues and ideas from their perspectives, from which we created a public forum, titled *"Social Media: The Good, the Bad and the Possible,"* which brought together the general public, digital and health care experts, dignitaries, and researchers to engage in dialogues on how social media can be used to improve their own health and wellness. The forum, which attracted an audience of over 195 live participants locally, over 100 participants via Web casting, 1004 tweets, and 1,124,886 impressions following the #HCSMForum hash tag on Twitter from BC and beyond, demonstrated the keen interest of the general public in this topic.

Figure 1. Participants of the Peter Wall Institute of Advanced Studies funded workshop "Harnessing the Social Web for Health and Wellness: Issues for Research and Knowledge Translation", in Vancouver, Canada, February 3 - 5, 2013.

Workshop Results

The success of this workshop was built on the intellectual diversity and collaboration of a group of multidisciplinary experts, enriched by the experiences of users of social media in the public. Our primary intention was to address a health audience and a social media audience of fellow researchers and academics in different disciplines who are interested in learning about and/or engaging in ways social media can be harnessed in the service of health and wellbeing. Our secondary audience included: (1) the general public to give us insights on the everyday use of social media and health practices to potentially participate with us in our research; (2) research institutions' decision makers to raise their interest in this area of research to support further exploration of this important sociotechnical phenomenon; and (3) knowledge users wanting to leverage this medium to help improve wellness of our citizens (health policy makers, businesses and corporations, engineering/computer science enterprises, social scientists, health professionals).

The pre-workshop preparations on the 6 questions, and the input and discussions at the research workshop and the public forum, served as the basis of the 3 manuscripts associated with this workshop published in JMIR [1-3]. Their content will not be detailed in this editorial, but readers are encouraged to access these papers. The discussions in the workshop brought out several key insights which crossed disciplinary boundaries, highlighted in Textbox 2. These are essential questions and issues that need to be further investigated and addressed through interdisciplinary research.

Textbox 2. Key multidisciplinary insights from discussion in the Peter Wall workshop.

- *What's the evidence:* While there was a common notion that social media can be very helpful to support health and wellness, how do we quantify these benefits? What evidence needs to be generated to support this impression?
- *The public:* How could social media optimally and effectively support the public's active pursuit of health and become experts of their own wellness? How could we improve the different segments of the public, including but not limited to age, language and technological literacy, types of social media used (Eg, twitter, facebook) in accessing the information and knowledge captured in the various dimensions of the social media?
- *Health professionals:* Health professionals are not embracing social media as rapidly as the public for health services or information dissemination. How could we incentivize health professionals to embrace social media, track its evolution, and work with the public to fully unleash the power of this medium to support health and wellness? Meanwhile, is there a professional obligation for health professionals on social media to respond to inquiries or address misinformation if they choose to participate in social media?
- *The information on social media:* Is the depth of the information or knowledge shared on social media appropriate for pursuit of health? How could the interactivity of the dialogues on social media be best used to support health and disease management? How to build trust and credibility to nurture the therapeutic relationships between health professionals and the patients via social media? How to increase involvement of individuals, and what types of roles are needed amongst the different participants to sustain social media discussions? Should these roles be assigned or could they be allowed to emerge spontaneously as the virtual community grows?
- *The evolution of social media:* The disruptive nature of social media to alter communal knowledge sharing is creating positive tension to stimulate changes. How can we harness this momentum of change towards empowering health consumers and health professionals alike in knowledge exchange and dissemination in health? What types of electronic tools within and outside of social media will continue to emerge in the future? How can individuals keep abreast of these changes and continue to adopt them for health and wellness?

Next Steps

Together, the workshop and public forum shed light on the need to engage the general public, health professionals, researchers, and innovators to work together to unleash the power of social media and modern information technologies. The workshop further shed light on new and exciting ways to leverage social media in health care and to improve our health system's ability and capacity to support each of us in wellness and sickness alike. In addition to disseminating the findings through publications, we were delighted to be engaged by the British Columbia Patient Safety and Quality Council to co-organize a follow up workshop, scheduled to occur in early 2014, to uncover the most beneficial ways of leveraging social media to promote quality and health in BC. Insights and lessons learned from this workshop will also be incorporated into the eHealth educational curriculum of medicine at UBC, and nationally through the Royal College of Physicians and Surgeons of Canada eHealth Expert Working Group and the Association of Faculties of Medicine of Canada eHealth education Committee.

We further intend to maintain connections with our workshop and public forum participants, and invite others interested in this endeavour to join and continue to explore this subject. For more information, please contact the corresponding author.

Acknowledgments

This workshop was funded by a grant from the Peter Wall Institute of Advanced Studies. The author would like to recognize the University of British Columbia co-investigators and leads from the different schools: Caroline Haythornthwaite from School of Library Science, Leanne Currie from the School of Nursing, Hasan Cavusoglu from the Sauder School of Business, and Sid Fels from Engineering and Helen Novak Lauscher from Faculty of Medicine eHealth Strategy Office for steering the development and implementation of this workshop. The author would like to thank the 30 participants at the workshop, Lindsay Zibrik from the eHealth Strategy Office for the administrative organization of the workshop, Francisco Grajales III and Ajay Puri for leading the social media strategy. The author would also like to thank the public forum moderator and panel speakers: Gillian Shaw from the Vancouver Sun newspaper Ana Lucia Ruggiero from Pan American Health Organization, Gunther Eysenbach from University of Toronto, Caroline Haythornthwaite, Helen Novak Lauscher, and Francisco Grajales III, and the general public audience members that participated in person or virtually in the forum. Finally, the author would like to acknowledge the Harnessing the Social Web Workshop Collaborators in the collective contributions to the content of this editorial.

List of Peter Wall workshop participants:

Richard Booth; Western University

Elizabeth Borycki; University of Victoria

Lorna Boschman; University of British Columbia

Joan Bottorff; University of British Columbia's Okanagan campus

Hasan Cavusoglu; University of British Columbia

Leanne Currie; University of British Columbia

Gunther Eysenbach; University of Toronto

Larry Frisch; Vancouver Coastal Health Research

Bernie Garrett; University of British Columbia

Jennifer A Gibson; University of British Columbia

Francisco Grajales III; University of British Columbia

Anatoliy Gruzd; Dalhousie University

Caroline Haythornthwaite; University of British Columbia

Kendall Ho; University of British Columbia

Christina Krause; BC Patient Safety and Quality Council

Andre Kushniruk; University of Victoria

Bernadette Mah; Peter Wall Institute for Advanced Studies

Sean A Munson; University of Washington

Cameron Norman; CENSE Research and Design

Helen Novak Lauscher; University of British Columbia

John Oliffe; University of British Columbia

Ajay Puri; BC Patient Safety and Quality Council

Lucia Ruggiero; Pan American Health Organization/World Health Organization

Jat Sandhu; Vancouver Coastal Health Authority; University of British Columbia

Sophia Woodman; University of British Columbia

Lindsay Zibrik; University of British Columbia

Conflicts of Interest

None declared.

References

1. Gruzd A, Haythornthwaite C. Enabling community through social media. J Med Internet Res 2013;15(10):e248 [FREE Full text] [doi: 10.2196/jmir.2796] [Medline: 24176835]
2. Munson SA, Cavusoglu H, Frisch L, Fels S. Sociotechnical challenges and progress in using social media for health. J Med Internet Res 2013;15(10):e226 [FREE Full text] [doi: 10.2196/jmir.2792] [Medline: 24148206]
3. Grajales IIIFJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social Media: A Review and Tutorial of Applications in Medicine and Health Care. J Med Internet Res 2014;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912]
4. Wikipedia. Social media URL: http://en.wikipedia.org/wiki/Social_media [accessed 2013-11-07] [WebCite Cache ID 6KwTmGGWn]
5. Deloitte Development LLC. 2010. Social networks in health care: Communication, collaboration and insights URL: http://bit.ly/17Qcv5v [accessed 2014-01-16] [WebCite Cache ID 6MfB9bIC2]
6. Haythornthwaite C. Crowds and communities: Light and heavyweight models of peer production. In: Proceedings of the 42nd Hawaii International Conference on System Sciences. 2009 Presented at: 42nd Hawaii International Conference on System Sciences; January 5 to 8, 2009; Waikoloa, Big Island, Hawaii.
7. Benkler Y. The wealth of networks: how social production transforms markets and freedom. New Haven [CT]: Yale University Press; 2006.
8. Ho K, Jarvis-Selinger S, Norman CD, Li LC, Olatunbosun T, Cressman C, et al. Electronic communities of practice: guidelines from a project. J Contin Educ Health Prof 2010;30(2):139-143. [doi: 10.1002/chp.20071] [Medline: 20564704]
9. Chretien KC, Kind T. Social media and clinical care: ethical, professional, and social implications. Circulation 2013 Apr 2;127(13):1413-1421. [doi: 10.1161/CIRCULATIONAHA.112.128017] [Medline: 23547180]
10. Chou WY, Prestin A, Lyons C, Wen KY. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]
11. Kling R, Rosenbaum H, Sawyer S. Understanding And Communicating Social Informatics: A Framework For Studying And Teaching The Human Contexts Of Information And Communication Technologies. Medford, NJ: Information Today; 2005.

12. Rainie L, Wellman B. Networked: The New Social Operating System. Cambridge, MA: The MIT Press; 2012.

Edited by G Eysenbach; submitted 19.09.13; peer-reviewed by A Djuricich, A Burke; comments to author 07.10.13; revised version received 09.11.13; accepted 12.01.14; published 11.02.14.

Please cite as:
Ho K, Peter Wall Workshop Participants
Harnessing the Social Web for Health and Wellness: Issues for Research and Knowledge Translation
J Med Internet Res 2014;16(2):e34
URL: http://www.jmir.org/2014/2/e34/
doi:10.2196/jmir.2969
PMID:24518432

Viewpoint

How Can Research Keep Up With eHealth? Ten Strategies for Increasing the Timeliness and Usefulness of eHealth Research

Timothy B Baker[1*], PhD; David H Gustafson[2*], PhD; Dhavan Shah[3*], PhD

[1]Center for Tobacco Research and Intervention, School of Medicine and Public Health, University of Wisconsin - Madison, Madison, WI, United States

[2]Center for Health Enhancement Systems Studies (CHESS), Industrial & Systems Engineering, University of Wisconsin - Madison, Madison, WI, United States

[3]Mass Communication Research Center, School of Journalism and Mass Communication, University of Wisconsin - Madison, Madison, WI, United States

[*]all authors contributed equally

Corresponding Author:
Timothy B Baker, PhD
Center for Tobacco Research and Intervention
School of Medicine and Public Health
University of Wisconsin - Madison
1930 Monroe St
Madison, WI, 53711
United States
Phone: 1 608 692 2009
Fax: 1 608 265 3102
Email: tbb@ctri.medicine.wisc.edu

Abstract

Background: eHealth interventions appear and change so quickly that they challenge the way we conduct research. By the time a randomized trial of a new intervention is published, technological improvements and clinical discoveries may make the intervention dated and unappealing. This and the spate of health-related apps and websites may lead consumers, patients, and caregivers to use interventions that lack evidence of efficacy.

Objective: This paper aims to offer strategies for increasing the speed and usefulness of eHealth research.

Methods: The paper describes two types of strategies based on the authors' own research and the research literature: those that improve the efficiency of eHealth research, and those that improve its quality.

Results: Efficiency strategies include: (1) think small: conduct small studies that can target discrete but significant questions and thereby speed knowledge acquisition; (2) use efficient designs: use such methods as fractional-factorial and quasi-experimental designs and surrogate endpoints, and experimentally modify and evaluate interventions and delivery systems already in use; (3) study universals: focus on timeless behavioral, psychological, and cognitive principles and systems; (4) anticipate the next big thing: listen to voices outside normal practice and connect different perspectives for new insights; (5) improve information delivery systems: researchers should apply their communications expertise to enhance inter-researcher communication, which could synergistically accelerate progress and capitalize upon the availability of "big data"; and (6) develop models, including mediators and moderators: valid models are remarkably generative, and tests of moderation and mediation should elucidate boundary conditions of effects and treatment mechanisms. Quality strategies include: (1) continuous quality improvement: researchers need to borrow engineering practices such as the continuous enhancement of interventions to incorporate clinical and technological progress; (2) help consumers identify quality: consumers, clinicians, and others all need to easily identify quality, suggesting the need to efficiently and publicly index intervention quality; (3) reduce the costs of care: concern with health care costs can drive intervention adoption and use and lead to novel intervention effects (eg, reduced falls in the elderly); and (4) deeply understand users: a rigorous evaluation of the consumer's needs is a key starting point for intervention development.

Conclusions: The challenges of distinguishing and distributing scientifically validated interventions are formidable. The strategies described are meant to spur discussion and further thinking, which are important, given the potential of eHealth interventions to help patients and families.

(J Med Internet Res 2014;16(2):e36) doi:10.2196/jmir.2925

KEYWORDS

social media; Internet; randomized clinical trials; experimental designs; research techniques; patient education; patient engagement; health communication; telemedicine

Introduction

eHealth interventions are appearing and changing so quickly that they are challenging the way we conduct health care research. For the purposes of this paper, we adopt a broad definition of eHealth in keeping with Eysenbach's definition [1]. That is, we view eHealth as an attempt to enhance health or health service delivery through use of modern information technology and electronic communication resources [2]. Thus, in our view, eHealth comprises interventions involving the Internet, wireless communications, interactive TV, voice response systems, kiosks, personal digital assistants (PDAs), CD-ROMs, DVD-ROMs, and remote monitoring that guides intervention delivery. Because eHealth interventions—for example, mHealth, telemedicine, information and communication systems—are defined in part by the technology they are built upon, their nature, relevance, appeal, and uniqueness are all affected by the rapid pace of technological change. By the time a randomized trial of a new intervention takes place, updated technology is likely to make the tested intervention and the results of the trial out-of-date. A recent paper [3] addressing the need to speed up the research enterprise noted that it often takes 7 years to submit a grant, design and pilot test the methods, conduct the research, analyze the data, and publish the results. If this span consisted of the years 2006-2012, the following innovations would have occurred in that period: the Wii, the iPhone, the Android system and products, the iPad, and Twitter. What have we learned about conducting research in such a fast-changing world and what can we do about it? The past gives us some guidance.

A Brief History

We have conducted randomized trials on eHealth interventions since the 1970s for a variety of chronic conditions, such as cancer, asthma, and addiction [4-8]. When we started exploring the effectiveness of eHealth interventions [5,6,9,10], iPads, iPods, smartphones, Twitter, Facebook, and even the Internet were things of the future. We began by creating interventions to be used on desktop computers.

When we started, we believed desktop computers and material displayed on monitors would be an important way to communicate for decades. This assumption guided our work on the services contained in two early systems we developed, BARN [10] (the Body Awareness Resource Network, intended for adolescents) and CHESS (Comprehensive Health Enhancement Support System) [7]. For example, the early versions of our CHESS breast cancer interventions contained services such as an "Instant Library", answers to "Frequently Asked Questions", "Personal Stories" of patients struggling with breast cancer, a treatment decision support system, a discussion group, and "Ask the Expert". Individually, these services were quite new and because they were integrated and coordinated within an invitation-only website (a "walled garden"), they were unique. Our randomized trials found that participants in our studies who had access to CHESS used it heavily and did better clinically than participants receiving usual care or who had unfettered access to the whole Internet but did not have access to CHESS [8,11,12]. In particular, we found that participants heavily used the "social" components of the system (ie, the discussion groups).

When we first started to test CHESS systems in the mid-1980s, we loaned participants desktop computers (Apple IIs) and arranged and paid for their dial-up connection (to either the Internet or to a computer-linked modem in pre-Internet days). This was the first time many participants had used a computer or the Internet. This no doubt motivated some participants to volunteer for the research and be active users of the new system [13-16].

Our early desktop/laptop-based interventions are now dated. We have trouble recruiting breast cancer patients into studies using desktop/laptop interventions, and once we recruit individuals into the research, participants do not use CHESS as much as they once did, especially the social resources. Moreover, desktop/laptop CHESS systems for breast cancer do not appear to confer the same benefits they once did [17].

PC-based CHESS systems for breast cancer were novel 20 years ago, but they now offer less functionality as a social resource than do websites such as Facebook. Some people may still prefer the "walled garden" of CHESS, which confers confidentiality and vets participants, but the ability to create "circles" within existing social networking platforms has lessened this distinction. In essence, about 10 years ago participants in our research found CHESS to be novel and appealing, and our research showed that it was heavily used and exerted large effects relative to access to other Internet resources. That has all changed substantially.

The speed of this transition is breathtaking and can be appreciated by comparing the shelf-life of CHESS with that of other sorts of psychosocial interventions: for example, with behavioral interventions that have existed in similar forms for the past 50 years and with psychodynamic psychotherapy, some forms of which are still being offered and used much as they were a century ago. Thus, in the past, researchers could afford to conduct research at a desultory pace; the results of their research would be relevant for many years. We believe that the history of CHESS is a harbinger of what is in store for virtually all eHealth intervention strategies. They will become dated and remarkably quickly. This will occur because a defining feature is the nature of their delivery mechanism(s)—which is vulnerable to the breakneck pace of technological change.

The Challenges for Research

eHealth interventions take time to evaluate in part because they are so complex. They involve a combination of content, user, social interaction, platform, links, and interface, making them intrinsically complicated to study. An intervention might have greater real-world utility to the extent that it permits adaptation in these dimensions, yet this same complexity and adaptability can complicate the evaluation of an intervention (and of course, too much complexity might hinder use and efficacy). Is a cancer intervention the same if delivered on highly divergent platforms with unique functionality (smartphone vs desktop)? What is being evaluated if the intervention also permits broad access to the Internet? Is it the intervention or Twitter, Facebook, WebMD, or the Cancer Survivor Network that is driving the observed effect? Such adaptability may obfuscate the nature of the intervention [18], produce great variation or error in the effects of the intervention, and, because these linked resources will themselves change over time, the evaluation itself will be built on shifting sand.

The dissemination of eHealth resources (or related applications such as social media) can occur very rapidly. Thus, a commercial developer can produce a resource and release it in short order so that the resource is in the hands of thousands of users, with nothing being known about its efficacy. Such interventions may become more widely used than validated interventions for reasons other than their effectiveness (eg, search engine status or appearance). In few fields is an intervention made available to so many, so quickly, as in eHealth, with some even going viral. Rapid dissemination over the Internet not only increases the rate at which new interventions enter the field but also the rate at which extant interventions are rendered out-of-date and unappealing. Hence another irony: An intervention that is tested and experimentally validated before dissemination may become less widely used because its content, functions, and platform are no longer innovative by the time it is disseminated. The corollary is, of course, that the widely used, novel, and untested intervention may be inert or even harmful and could displace the use of effective interventions.

Even in the best of circumstances, when an intervention has been shown to be effective and intensively used, it will be copied by numerous competitors so that the research version of the intervention is ubiquitous, seems commonplace, and cannot be cleanly evaluated because individuals in control conditions have ready access to its components through competing systems. It would be hard to evaluate psychoanalysis if every friend or neighbor processed transference, did dream interpretation, and maintained therapeutic neutrality.

A final challenge to conducting research on eHealth interventions is the pace of *clinical* discoveries. When we first produced the CHESS breast cancer intervention, we made wholesale updates of the content annually. This periodicity was fine in the 1980s and 1990s, but faster updating must be done now as a consequence of accelerating progress in fields such as the radiology, oncology, and pharmacogenomics [19]. Efficient strategies can be used to maintain some currency, such as including links to current literature that can be updated quickly and cheaply. However, the core of the information presented by the intervention and its integration with such features as treatment decision-making need to be current or the intervention loses value and credibility among clinicians and patients. Finding and hiring highly skilled experts to help with the updating is difficult and expensive, and even a regularly updated eHealth intervention may have a short life. Again, we can evaluate a given instantiation of the intervention (including on-going attempts to keep it current), but its overall novelty and appeal can erode over the course of a study that may last years and will certainly change in dissemination.

Increasing the Efficiency of eHealth Research

The rapid changes occurring in technology and the pace of medical research highlight a need for eHealth research strategies that both increase the pace of research and also produce higher-value interventions that will be more effective than the ascientific application of technology (eg, by app developers) and therefore, remain appealing and effective despite tumultuous change. In other words, researchers can improve eHealth by enhancing both the efficiency and rapidity of eHealth research and its quality and merit. We will start by suggesting ways to accelerate the pace of eHealth research and then consider how to enhance the quality of interventions.

Think Small: Towards More Focused and Efficient Research Studies

Large clinical trials designed to evaluate the effects of whole eHealth interventions may take many years, considering the time needed to secure funding, conduct recruitment and implementation, and so on [3]. However, researchers can conduct small-N studies that efficiently target relatively discrete questions. This research can occur either in the laboratory setting (eg, if special equipment is needed such as eye-tracking or physiological apparati) or with small samples of real-world users (eg, if opportunities exist for examining existing systems). Many questions might be addressed effectively with small sample ad hoc experiments, which may allow many questions to be addressed simultaneously. Which tailoring features do people prefer? Which methods of framing information help individuals remember key points? The key is to detach addressing such questions from the analysis of an entire eHealth intervention that occurs in the context of a large and slow-to-complete clinical trial [20].

Certain strategies go hand-in-hand with this smaller, focused approach, beginning with using more proximal outcome measures, ones that are clinically meaningful but also highly sensitive and quickly responsive to the effects being evaluated [21]. One reason that AIDS research may have progressed rapidly compared with cancer research is its focus on “surrogate endpoints”, such as viral load, that are highly sensitive to therapeutic change and that have reduced reliance on distal clinical outcomes, such as survival duration [22]. This sort of focus might lead eHealth researchers toward using meaningful yet efficient outcomes such as self-efficacy, increased medication adherence, and greater perceived social support. Such outcomes could not substitute for vital clinical endpoints, but they could be used in ongoing research that results in continuous improvement [23].

Second, some experimental designs are geared for small sample studies. These designs are referred to with various labels such as “single-subject”, “multiple baseline”, “stepped wedge”, and “quasi-experimental” designs. These involve the systematic presentation and/or removal of interventions or intervention components from participants and determining whether meaningful change occurs contingent upon such manipulations [24,25]. Importantly, new analytic methods are appearing that increase the internal validity of such efficient strategies (eg, de Vries & Morey using Bayes’ tests for single-subject data [26]).

Use Efficient Research Designs

Evaluation of eHealth treatments often occurs via randomized clinical trials. While there is a vital role for such trials, they often do not provide as much information as alternative experimental strategies. For instance, engineering researchers [23] typically use highly efficient factorial and fractional-factorial designs that allow for the testing of multiple hypotheses or interventions with no loss of power even as the number of tested interventions increases. Collins et al note that testing 6 intervention features or components would require 6 different studies if traditional randomized controlled trial (RCT) designs were employed (comparing an active component with a control/placebo component in each study) [23]. However, a single experiment could contrast all 6 intervention components if they were tested in a 6-factor factorial design (with each factor comprising an active and control component) with every participant being independently randomly assigned to each factor.

Factorial designs have some decided advantages over the traditional RCT approach. For example, the 6-factor factorial design is far more efficient; for example, using the same targeted effect size, the factorial experiment would have the same power to test each factor as would each RCT and use about one-sixth the participants. In addition, the factorial experiments would allow the investigator to estimate interaction effects among intervention components, which would indicate which combinations of intervention components worked best together—something not possible (efficiently) with conventional RCTs. And certainly the investigator could conduct the factorial experiment in less time than it would take to do 6 RCTs.

Other efficient research strategies or designs might also speed the research process, namely sequential, multiple assignment, randomized trials (SMART) or adaptive designs [27-29]. Such designs are appropriate for conditions or problems where a change in a patient’s status might require a change in treatment approach; for example, smoking treatment might be changed when a patient trying to quit relapses back to regular smoking. In SMART designs, the researcher may not only investigate multiple intervention components in the same study, but do so as participants transition across the various phases of recovery [20]. Further, one can also ask research questions that rely not on new interventions, but on systems already in use. This could reduce development time and costs and speed the dissemination of findings. One of the authors has recently modified the existing National Cancer Institute smoke-free website by recruiting smokers who visit the site and randomizing them to different website versions. With such a strategy, researchers can simultaneously evaluate reach, effectiveness, and maintenance [30] (see Riley for additional strategies to speed the research enterprise [3]). This strategy of using existing interventions and delivery systems seems especially appropriate when ambitious, comprehensive research questions are examined—for instance, those involving multiple aspects of effectiveness (eg, across communication, control, care, and contextual dimensions [20]). In this way, the time needed for development and implementation are not added to the time needed for evaluation.

Study Universals

Research evaluating eHealth interventions often addresses the effects of a particular intervention, which can delimit the relevance of the research (eg, as the intervention becomes dated, so do the results of the research). To increase the odds that research yields durable and broadly relevant results, research could examine timeless behavioral, psychological, and cognitive principles and systems. These principles and systems are typically generated by well- supported basic theory (eg, theories of behavior change or quality of life, such as self-determination theory or other broad social science theories; see Kaplan et al [20] and Niemiec et al [31]). It is also possible, though, that theory that is more applied in nature could also reveal universal principles (eg, theories of general system design and others, such as Dansky [32] and Yen & Bakken [33]). What are the principles by which information is made more salient? How should people be queried to help them arrive at optimal decisions? What general approaches increase motivation in an eHealth context? What sorts of messages most efficiently provide emotional and instrumental support? Just as psychometricians validate an assessment instrument across multiple populations to reduce sampling error, eHealth researchers should validate principles across diverse interventions and platforms (eg, with efficient factorial designs), thus building into research the demonstration of broad relevance. Of course, such research should search for moderation effects to determine just how “universal” the phenomena or principles are. For instance, it may be the case that some technology-intensive interventions will be inappropriate for developing countries [34].

Anticipate the Next Big Thing

Researchers need to anticipate eHealth strategies that will work in future environments, not in the present [35,36]. For instance, to figure out new approaches to what we regarded as an unsustainable addiction treatment system, we at the University of Wisconsin Center for Health Enhancement Systems Studies convened a meeting of drug addicts, family members, biomedical engineers, nano-technologists, futurists, computer scientists, and experts in social networking, quality improvement, and pharmacology. Only two people were from addiction treatment. The addicts and families told their stories. A futurist reviewed where the world was likely to go generally. The group was told that a virus had selectively killed every addiction treatment provider. The attendees had to design an addiction treatment system built solely on technology. Attendees broke into groups, each containing an addict, a family member, and experts. Many new insights emerged, including those that led to our work in smartphones and sensors [37]. The key to our innovation in this area was assuming that we could depart entirely from prior approaches to addiction treatment and to enlist very new outside perspectives in conceptualizing change.

The "next big thing" also certainly involves incorporating the latest technology into eHealth interventions. Using mobile devices has enabled us to create services we could not have imagined 20 years ago. We can build into CHESS features that take advantage of standard functions of smartphones, such as accelerometers, GPS, two-way video cameras, and magnetometers. These features allow us to create what are being called "ecological momentary assessments and interventions". For example, the smartphone-based A-CHESS (Addiction—Comprehensive Health Enhancement Support system) includes a service to track the movement of people recovering from alcoholism. If a user gets near a bar he or she used to frequent, the GPS initiates a rescue service by sending alerts to the participant and making calls to pre-designated friends or family [38]. Some participants find it easier to listen to content than read it; on the smartphone, most content can be presented auditorily. Other sensors enable us to measure almost innumerable mental and physical capabilities, creating the opportunity to use the phone's features to identify physical and mental indicators of stress and dysfunction and automatically request help. Using the latest technology and listening to new viewpoints in designing the A-CHESS intervention seem to be paying off; early results of this intervention have been very encouraging [39].

Improve Information Delivery Systems

Even when researchers make important and timely discoveries, it is difficult and time-consuming to disseminate them to other researchers. Traditional dissemination vehicles such as conferences do not keep pace with rapid developments in research. A mantra of our research center is that no one should suffer twice. Dealing with an injury or disease is tough enough. Doing so in an inefficient, complex delivery system adds suffering. So, certainly attention to methods for improving communication and "handoffs" among providers, patients, and family should be a key goal of eHealth research [40-43]. Handoffs of results between researchers should similarly be improved by new, innovative methods—methods that apply the same conceptual analysis, effort, and information technology resources that are used in the designs of their eHealth interventions. Specifically, more research and effort need to be expended on vehicles and resources that allow eHealth researchers to communicate efficiently with one another and share research experiences, products, and resources (eg, prepublication findings, solutions to technical problems, coding resources, and so on). Such resources could support the development of research teams that share intervention development burdens and jointly recruit for studies and could foster the use of common measures that would promote "big data" research (see the section below, "Develop models, including mediators and moderators"). These steps could significantly accelerate the conduct of individual studies and the overall pace of research as well. Barriers to such developments certainly exist, such as concerns about authorship credit for developing interventions and reporting findings and the work and money needed to maintain such resources. But these barriers could be overcome (eg, by allowing website posting to constitute a claim to authorship). These sorts of resources would not compete with normal channels of research communication (eg, peer-reviewed journals) but would complement the research that appears in such outlets. This would require expansion of the sort of Web resources that have been developed for other research domains (eg, the research methodology website sponsored by Pennsylvania State University [44]). Certainly the communication and technological sophistication of eHealth researchers could be leveraged to address their own communication problems.

Develop Models, Including Mediators and Moderators

While vast amounts of data exist from eHealth experiments, the data have not been mined systematically. eHealth has generated extensive databases. Enormous amounts of information lie within keystrokes and in messages, posts, and chats and can—with users' permission—be analyzed to develop decision-support systems that help users address their concerns more efficiently and effectively. Mathematical modeling and simulation can help transform data into information. Bayesian models have been developed to predict whether a person will make an attempt on his or her life [5]. Simulations can rapidly compare treatment alternatives. Where data do not exist, methods for quantifying expert judgments can be employed [45]. The availability of such databases permits the evaluation of important research questions without developing interventions and implementing them or conducting new clinical trials. Such "big data" approaches to research would echo development in other areas of research such as dBGap, a repository for GWAS (genome-wide association study data) and related phenotype data. The researcher communication Web resources described above could similarly house eHealth datasets.

Large existing databases would be an ideal resource for conducting tests of mediation and moderation. Mediation analyses suggest how interventions work, and moderation analyses identify factors that modulate how well interventions work. Moderation analyses, when done with pre-existing datasets, would dramatically reduce the time spent in obtaining "generalizability data" [46], allowing researchers to discover facets across which findings can be generalized: In which persons and contexts does an intervention work well? These questions lie at the heart of most comprehensive eHealth evaluation models [20,32,33]. Researchers should also conduct mediational analyses across different contexts to find general mechanisms of effect [47]. Mediation research is important because it can tell us if our theory of the intervention is correct. Is it working as hypothesized? Discovering how right, or wrong, our theories are could save untold time that might be spent going down blind alleys. Moreover, mediation analysis can tell us not only what an intervention is doing, but also what it is *not* doing (eg, not increasing knowledge of treatment side effects, not improving affect). Such information is vital to efficiently revising both our theories and our interventions, and it can now be more efficiently implemented with the development of multiple-mediator analytic strategies [48].

Improving the Quality of eHealth Interventions

Continuous Quality Improvement

Possibly because researchers know that countless studies can be done to assess the effects of a particular intervention, they are reluctant to change an intervention before it has been thoroughly evaluated. This, of course, produces a stagnant island in a sea of change. The recent biography of Steve Jobs [49] relates that Jobs and his colleagues at Apple were often not the first to think of or develop a product. Various portable music players preceded the iPod, tablets pre-existed the iPad, and so on. But Apple made the product better than anyone else—for instance, easier to use and more elegant. Sometimes the product did not start out better, but became better through rapid quality improvement (the original iPad did not have a camera). Perhaps eHealth researchers need to think more like engineers, who tend to use the continuous improvement principle whereby every product is in a sense a beta version, that is, always the target of improvement and refinement, as in Collins [23]. This meshes with recent calls to conduct ongoing, systematic assessment of eHealth interventions across their life cycles [2,33]—assessment that taps diverse intervention dimensions and effects (eg, quality of use, impact on workflow, costs). While called for, such comprehensive iterative evaluation certainly remains rare [33].

If an eHealth intervention is repeatedly altered for purposes of quality improvement, how can its clinical effects be evaluated? Can one evaluate an intervention that never “stays still”? Actually, this is possible. One strategy would be to compare the intervention longitudinally through its various improvements against some reasonable control condition such as “adlib” Internet access, which itself would be changing over time. Advances in intensive longitudinal modeling [50] permit powerful and focused tests in such multiphased longitudinal datasets.

Help Consumers Identify Quality

New eHealth developments are appearing rapidly, but “consumers” have very few ways to identify valuable ones (and consumers could include health care systems, clinicians, and others, in addition to patients and patients’ support networks). Many apps and eHealth interventions appear to have substantial weaknesses [51]. To complicate matters, eHealth systems are ever changing. Hence a one-time evaluation of an intervention can have limited value. Even identifying what new products exist can be very difficult. Consumers, like researchers, are challenged to keep up with such rapid development. New ways are needed to help consumers (and funders) make informed decisions about products. However, multiple barriers exist to addressing this need. For instance, rating eHealth quality comprehensively can be complex and difficult [51], especially if evaluations target product effectiveness or multiple intervention dimensions [52]. A more feasible approach might be to start by evaluating eHealth resources systematically with easy-to-assess criteria, for example, “transparency criteria” [51], such as disclosure of authorship, sponsorship, and/or ownership; recency of last update; authors’ credentials; nature of the review process for information accuracy; and so on. Other fairly brief rating systems such as DISCERN [53] might also be modified to provide relevant rating dimensions [54]. However, this still leaves unanswered questions such as who would conduct such ratings and how the ratings would be promulgated. These challenges may not be too daunting. For instance, a consortium of research organizations with some modest government sponsorship should be able to conduct ratings of numerous websites efficiently and cost-effectively. Assuming the use of an easy-to-use rating instrument, such ratings would be far less arduous and costly than Cochrane evaluations. Moreover, the promulgation and advertisements of such ratings would seem straightforward—all rated and approved websites could prominently display evidence of their meeting the quality rating criteria, and their communications could educate the public to turn to eHealth resources that meet and display such approval. In other words, the dissemination medium would be the rated websites themselves. It is important to note that once a mechanism for the relatively basic evaluation of eHealth resources is developed, this system could be used to support more ambitious evaluations of quality, such as those addressing evidence of accuracy, completeness, reading level, design, and effectiveness [51].

Reduce Costs of Care

Reducing costs is often thought to be the enemy of quality, yet from numerous perspectives (societal, health care system), demonstrating cost reduction is very important and perhaps the chief means of increasing the dissemination of eHealth interventions (because it will appeal to decision-makers and purchasers). Yet, cost effectiveness or benefit is too infrequently demonstrated or identified as a key goal of research. For instance, reviews of available studies either report very little evidence on cost savings or specifically cite this as a lack in the field [55-58]. The impact of eHealth on costs will drive health system decisions for years to come. Properly designed, eHealth programs hold promise for reducing costs by speeding recovery and reducing admissions. eHealth might reduce the costs of care in many ways. For instance, informational resources might (1) reduce the frequency of medical staff contacts, (2) facilitate communication in chronic care intervention teams so that care is efficiently shifted to lower-cost providers, (3) directly deliver psychosocial interventions, thereby reducing the use of professional care, (4) improve patient preparation for health care visits, making them more efficient and reducing the need for repeat visits, and (5) improve patient satisfaction with care (eg, by increasing perceived connection with caregivers) and thereby reduce health care plan churn. Future research on eHealth interventions should explicitly consider the cost-effectiveness and cost-benefit impacts of an intervention (including formulating a business plan [59]), and, when possible, incorporate measures to address these outcomes.

Deeply Understand Users

The rapid increase in new technologies raises the potential of innovation bias [60]—that is, developers becoming so infatuated with an innovation that user needs become secondary. Researchers and developers should deeply understand user assets and needs and how technologies can build on strengths to meet user needs. And it is important to recognize that users of the technology could be considered not only to be the people actually interacting with a device, such as patients and clinicians [61], but also closely linked others who are significantly affected by the device (eg, family members trying to provide care). When we began our University of Wisconsin Center of Excellence in Active Aging, the technical team (including programmers) visited the homes of frail elderly individuals, ate with elders at congregate eating facilities, and volunteered 4 hours a week at a senior center. This helped transform a job into a calling. It also helped reveal needs and assets that elders themselves did not recognize, which suggested innovative solutions. In one home, we observed an elderly person try to move without a walker and almost falling as a result. This taught us the danger for some of unaided movements of even a short distance. In response, we constructed the "screaming walker", which has a radio-frequency identification chip that alerts the person when he or she tries to step away unaided. Because the in-depth assessment of user needs and assets is time consuming, developers often limit investment in this crucial activity. Fortunately, evidence shows that eHealth developers are increasingly making user input and needs assessment key elements in intervention design and practices that continue across intervention development [61].

Discussion

Conclusions

This paper has attempted to identify reasons it is so difficult to evaluate the effectiveness of eHealth interventions so that evaluation results are used, relevant, and timely. We note that eHealth interventions themselves, and evaluations of them, have relatively short shelf-lives because of the pace of technological advances, the pace of medical-health care advances, and the production of Internet and electronic resources (eg, apps) by multiple, nontraditional intervention developers, which has increased the number and diversity of intervention types. Such changes may make it difficult to distinguish scientifically validated interventions. Moreover, research on quality and evidence of the effectiveness of eHealth interventions may be anachronistic by the time they are produced and difficult to disseminate to relevant audiences.

This paper has suggested directions that eHealth developers and researchers might take to develop and evaluate eHealth interventions so that their interventions are scientifically grounded, innovative, and attractive enough to be competitive in the eHealth marketplace. This paper also provides guidance for enhancing intervention quality and making information on quality more available to potential users. At this point, these recommendations are designed to spur discussion and further thinking; they are too imprecise and aspirational to constitute a blueprint for change. Also, obstacles stand in the way of their full pursuit and implementation. For instance, some are intrinsically very difficult to achieve (eg, creating highly innovative interventions that anticipate the next wave of technology), and others would be difficult to achieve because they require contributions from multiple stakeholders (eg, creating a consortium of developers, researchers, and others who would grade eHealth intervention quality and promulgate the ratings). However, the first step in overcoming a challenge is to recognize that it exists and then generate ideas to overcome it. The current paper is merely one attempt to encourage progress, especially "quality progress", in the field eHealth intervention and evaluation.

Continued research on eHealth is critical because such interventions have tremendous potential to provide many patients and families readily available and inexpensive assistance. However, investigators (including ourselves) need to conduct research that recognizes the rapid, ever-changing landscape of technological, scientific, and e-social progress. The notion that technological and scientific progress creates unanticipated change and may render eHealth knowledge and products anachronistic is not new. Unless researchers can discover ways to produce appealing and effective interventions that compete well in the eHealth marketplace, many individuals may use eHealth and mHealth resources that exert negligible or even iatrogenic effects.

Acknowledgments

The writing of this paper was supported by agencies of the National Institutes of Health: the National Cancer Institute under award numbers P50CA095817 and P50CA143188-11, and the National Institute on Alcohol Abuse and Alcoholism under award number R01AA017192. The funders had no role in preparing or reviewing the manuscript or deciding where to submit it. The authors would like to thank Bobbie Johnson and Wendy Theobald for invaluable assistance with the preparation of this paper.

Conflicts of Interest

None declared.

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. 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]

3. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013;2(1):10 [FREE Full text] [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]
4. Greist JH, Gustafson DH, Stauss FF, Rowse GL, Laughren TP, Chiles JA. A computer interview for suicide-risk prediction. Am J Psychiatry 1973 Dec;130(12):1327-1332. [Medline: 4585280]
5. Gustafson DH, Greist JH, Stauss FF, Erdman H, Laughren T. A probabilistic system for identifying suicide attemptors. Comput Biomed Res 1977 Apr;10(2):83-89. [Medline: 322942]
6. Gustafson DH, Bosworth K, Chewning B, Hawkins RP. Computer-based health promotion: combining technological advances with problem-solving techniques to effect successful health behavior changes. Annu Rev Public Health 1987;8:387-415. [doi: 10.1146/annurev.pu.08.050187.002131] [Medline: 3555528]
7. Gustafson D, Wise M, McTavish F, Taylor JO, Wolberg W, Stewart J, et al. Development and Pilot Evaluation of a Computer-Based Support System for Women with Breast Cancer. Journal of Psychosocial Oncology 1994 Mar 24;11(4):69-93. [doi: 10.1300/J077V11N04_05]
8. Gustafson DH, Julesberg K, Stengle W, McTavish FM, Hawkins RP. Assessing costs and outcomes of providing computer support to underserved women with breast cancer: a work in progress. Electronic J Commun (3-4) 2001;11:3-4.
9. Gustafson DH, Tianen B, Greist JH. A computer-based system for identifying suicide attemptors. Comput Biomed Res 1981 Apr;14(2):144-157. [Medline: 7273716]
10. Bosworth K, Chewning B, Day T, Hawkins R, Gustafson D. Barny a Computer Based Health Information System for Adolescents. The Journal of Early Adolescence 1981 Aug 01;1(3):315-321. [doi: 10.1177/027243168100100311]
11. 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 [FREE Full text] [doi: 10.1111/j.1460-2466.2008.00383.x] [Medline: 21804645]
12. 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 [FREE Full text] [Medline: 11520380]
13. Gustafson DH, Hawkins R, Boberg E, Pingree S, Serlin RE, Graziano F, et al. Impact of a patient-centered, computer-based health information/support system. Am J Prev Med 1999 Jan;16(1):1-9. [Medline: 9894548]
14. Gustafson DH, Hawkins RP, Boberg EW, McTavish F, Owens B, Wise M, et al. CHESS: 10 years of research and development in consumer health informatics for broad populations, including the underserved. Int J Med Inform 2002 Nov 12;65(3):169-177. [Medline: 12414016]
15. McTavish FM, Gustafson DH, Owens BH, Hawkins RP, Pingree S, Wise M, et al. CHESS (Comprehensive Health Enhancement Support System): an interactive computer system for women with breast cancer piloted with an underserved population. J Ambul Care Manage 1995 Jul;18(3):35-41. [Medline: 10143478]
16. Pingree S, Hawkins RP, Gustafson DH, Boberg E, Bricker E, Wise M, et al. Will the disadvantaged ride the information highway? Hopeful answers from a computer ‐ based health crisis system. Journal of Broadcasting & Electronic Media 1996 Jun;40(3):331-353. [doi: 10.1080/08838159609364357]
17. Baker TB, Hawkins R, Pingree S, Roberts LJ, McDowell HE, Shaw BR, et al. Optimizing eHealth breast cancer interventions: which types of eHealth services are effective? Transl Behav Med 2011 Mar;1(1):134-145 [FREE Full text] [Medline: 21709810]
18. Ahern DK, Patrick K, Phalen JM, Neiley JD. An introduction to methodological challenges in the evaluation of eHealth research: Perspectives from the Health e-Technologies Initiative. Eval Program Plann 2006 Nov;29(4):386-389. [doi: 10.1016/j.evalprogplan.2006.08.008] [Medline: 17950866]
19. Ingle JN. Pharmacogenomics of endocrine therapy in breast cancer. J Hum Genet 2013 Jun;58(6):306-312. [doi: 10.1038/jhg.2013.35] [Medline: 23635953]
20. 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]
21. Baker TB, Mermelstein R, Collins LM, Piper ME, Jorenby DE, Smith SS, et al. New methods for tobacco dependence treatment research. Ann Behav Med 2011 Apr;41(2):192-207 [FREE Full text] [doi: 10.1007/s12160-010-9252-y] [Medline: 21128037]
22. Thornton M. The next front in the war on cancer: faster clinical trials are critical if we are to save more lives. Wall Street Journal 2009 Feb 27:17.
23. 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 [FREE Full text] [doi: 10.1007/s12160-010-9253-x] [Medline: 21132416]
24. Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. Med Care 2011 Aug;49(8):761-768. [doi: 10.1097/MLR.0b013e318215d90d] [Medline: 21478771]
25. Kratochwill TR, Levin JR. Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. Psychol Methods 2010 Jun;15(2):124-144. [doi: 10.1037/a0017736] [Medline: 20515235]
26. de Vries RM, Morey RD. Bayesian hypothesis testing for single-subject designs. Psychol Methods 2013 Jun;18(2):165-185. [doi: 10.1037/a0031037] [Medline: 23458719]

27. Chow SC, Chang M. Adaptive design methods in clinical trials - a review. Orphanet J Rare Dis 2008;3:11 [FREE Full text] [doi: 10.1186/1750-1172-3-11] [Medline: 18454853]
28. Murphy SA, Lynch KG, Oslin D, McKay JR, TenHave T. Developing adaptive treatment strategies in substance abuse research. Drug Alcohol Depend 2007 May;88 Suppl 2:S24-S30 [FREE Full text] [doi: 10.1016/j.drugalcdep.2006.09.008] [Medline: 17056207]
29. Rivera DE, Pew MD, Collins LM. Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction. Drug Alcohol Depend 2007 May;88 Suppl 2:S31-S40 [FREE Full text] [doi: 10.1016/j.drugalcdep.2006.10.020] [Medline: 17169503]
30. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health 2003 Aug;93(8):1261-1267. [Medline: 12893608]
31. Niemiec CP, Ryan RM, Deci EL. Self-determination theorythe relation of autonomy to self-regulatory processespersonality development. In: Hoyle RH, editor. Handbook of personality and self-regulation. Malden, MA: Wiley-Blackwell; 2010:169-191.
32. 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]
33. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
34. van Gemert-Pijnen JE, Wynchank S, Covvey HD, Ossebaard HC. Improving the credibility of electronic health technologies. Bull World Health Organ 2012 May 1;90(5):323-323A [FREE Full text] [doi: 10.2471/BLT.11.099804] [Medline: 22589561]
35. Goldberg L, Lide B, Lowry S, Massett HA, O'Connell T, Preece J, et al. Usability and accessibility in consumer health informatics current trends and future challenges. Am J Prev Med 2011 May;40(5 Suppl 2):S187-S197. [doi: 10.1016/j.amepre.2011.01.009] [Medline: 21521594]
36. 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]
37. Gustafson DH, Palesh TE, Picard RW, Plsek PE, Maher L, Capoccia VA. Automating addiction treatment: enhancing the human experiencecreating a fix for the future. In: Future of Intelligent and Extelligent Health Environment: Volume 118 Studies in Health Technology and Informatics. Amsterdam: IOS Press; 2005:186-206.
38. Gustafson DH, Shaw BR, Isham A, Baker T, Boyle MG, Levy M. Explicating an evidence-based, theoretically informed, mobile technology-based system to improve outcomes for people in recovery for alcohol dependence. Subst Use Misuse 2011;46(1):96-111 [FREE Full text] [doi: 10.3109/10826084.2011.521413] [Medline: 21190410]
39. Gustafson DH, McTavish FM, Chih MY, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized controlled trial. JAMA Psychiatry 2014:- (forthcoming).
40. Ouwens M, Wollersheim H, Hermens R, Hulscher M, Grol R. Integrated care programmes for chronically ill patients: a review of systematic reviews. Int J Qual Health Care 2005 Apr;17(2):141-146 [FREE Full text] [doi: 10.1093/intqhc/mzi016] [Medline: 15665066]
41. Lilly CM, De Meo DL, Sonna LA, Haley KJ, Massaro AF, Wallace RF, et al. An intensive communication intervention for the critically ill. Am J Med 2000 Oct 15;109(6):469-475. [Medline: 11042236]
42. Coleman EA, Parry C, Chalmers S, Min SJ. The care transitions intervention: results of a randomized controlled trial. Arch Intern Med 2006 Sep 25;166(17):1822-1828. [doi: 10.1001/archinte.166.17.1822] [Medline: 17000937]
43. Van Eaton E. Handoff improvement: we need to understand what we are trying to fix. Jt Comm J Qual Patient Saf 2010 Feb;36(2):51. [Medline: 20180436]
44. The Methodology Center. Pennsylvania State University URL: http://methodology.psu.edu/ [accessed 2013-12-10] [WebCite Cache ID 6LlWa6PKn]
45. Gustafson DH, Fryback DG, Rose JH, Yick V, Prokop CT, Detmer DE, et al. A decision theoretic methodology for severity index development. Med Decis Making 1986;6(1):27-35. [Medline: 3080651]
46. Shavelson RJ, Webb NM. Generalizability Theory: A Primer. Thousand Oaks, CA: Sage; 1991.
47. Gustafson D, Wise M, Bhattacharya A, Pulvermacher A, Shanovich K, Phillips B, et al. The effects of combining Web-based eHealth with telephone nurse case management for pediatric asthma control: a randomized controlled trial. J Med Internet Res 2012;14(4):e101 [FREE Full text] [doi: 10.2196/jmir.1964] [Medline: 22835804]
48. Bolt DM, Piper ME, Theobald WE, Baker TB. Why two smoking cessation agents work better than one: role of craving suppression. J Consult Clin Psychol 2012 Feb;80(1):54-65 [FREE Full text] [doi: 10.1037/a0026366] [Medline: 22103958]
49. Isaacson W. Steve Jobs. New York, NY: Simon & Schuster; 2011.
50. a A. In: Walls TA, Schafer JL, editors. Models for intensive longitudinal data. Oxford, UK: Oxford University Press; 2006.
51. 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. [Medline: 12020305]
52. O'Grady L, Witteman H, Bender JL, Urowitz S, Wiljer D, Jadad AR. Measuring the impact of a moving target: towards a dynamic framework for evaluating collaborative adaptive interactive technologies. J Med Internet Res 2009;11(2):e20 [FREE Full text] [doi: 10.2196/jmir.1058] [Medline: 19632973]

53. DISCERN Online. University of Oxford URL: http://www.discern.org.uk/ [accessed 2013-12-10] [WebCite Cache ID 6LlWSe8nr]
54. Khazaal Y, Chatton A, Cochand S, Coquard O, Fernandez S, Khan R, et al. Brief DISCERN, six questions for the evaluation of evidence-based content of health-related websites. Patient Educ Couns 2009 Oct;77(1):33-37. [doi: 10.1016/j.pec.2009.02.016] [Medline: 19372023]
55. Donaldson LJ. From black bag to black box: will computers improve the NHS? BMJ 1996 Jun 1;312(7043):1371-1372 [FREE Full text] [Medline: 8646079]
56. 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. In: Evidence Report/Technology Assessment No. 175, AHRQ Publication No. 09-E004. Rockville, MD: Oregon Evidence-based Practice Center (Agency for Healthcare Research and Quality); 2008.
57. 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]
58. Lock C. What value do computers provide to NHS hospitals? BMJ 1996 Jun 1;312(7043):1407-1410 [FREE Full text] [Medline: 8646102]
59. van Limburg M, van Gemert-Pijnen JE, Nijland N, Ossebaard HC, Hendrix RM, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. J Med Internet Res 2011;13(4):e124 [FREE Full text] [doi: 10.2196/jmir.1674] [Medline: 22204896]
60. Rogers EM. Diffusion of innovations. 5th Edition. New York: The Free Press; 2003.
61. Whitehouse SR, Lam PY, Balka E, McLellan S, Deevska M, Penn D, et al. Co-Creation With TickiT: Designing and Evaluating a Clinical eHealth Platform for Youth. JMIR Res Protoc 2013;2(2):e42 [FREE Full text] [doi: 10.2196/resprot.2865] [Medline: 24140595]

Abbreviations

A-CHESS: Addiction—Comprehensive Health Enhancement Support System
CHESS: Comprehensive Health Enhancement Support System
eHealth: electronic health
e-social: social networking conducted electronically
mHealth: mobile health
RCT: randomized controlled trial
SMART: sequential, multiple assignment, randomized trials

Edited by G Eysenbach; submitted 29.08.13; peer-reviewed by R Glasgow, M Solomon; comments to author 30.10.13; revised version received 16.12.13; accepted 09.01.14; published 19.02.14.

Please cite as:
Baker TB, Gustafson DH, Shah D
How Can Research Keep Up With eHealth? Ten Strategies for Increasing the Timeliness and Usefulness of eHealth Research
J Med Internet Res 2014;16(2):e36
URL: http://www.jmir.org/2014/2/e36/
doi:10.2196/jmir.2925
PMID:24554442

Original Paper

Designing Messaging to Engage Patients in an Online Suicide Prevention Intervention: Survey Results From Patients With Current Suicidal Ideation

Ursula Whiteside[1], PhD; Anita Lungu[1], PhD; Julie Richards[1], MPH; Gregory E Simon[1], MD, MPH; Sarah Clingan[1], BA; Jaeden Siler[1], AA; Lorilei Snyder[1]; Evette Ludman[1], PhD

Group Health Research Institute, Seattle, WA, United States

Corresponding Author:
Ursula Whiteside, PhD
Group Health Research Institute
1730 Minor Ave Suite 1500
Seattle, WA,
United States
Phone: 1 206 287 2100
Fax: 1 206 287 2871
Email: whiteside.u@ghc.org

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2015/4/e69/

Abstract

Background: Computerized, Internet-delivered interventions can be efficacious; however, uptake and maintaining sustained client engagement are still big challenges. We see the development of effective engagement strategies as the next frontier in online health interventions, an area where much creative research has begun. We also argue that for engagement strategies to accomplish their purpose with novel targeted populations, they need to be tailored to such populations (ie, content is designed with the target population in mind). User-centered design frameworks provide a theoretical foundation for increasing user engagement and uptake by including users in development. However, deciding how to implement this approach to enage users in mental health intervention development is challenging.

Objective: The aim of this study was to get user input and feedback on acceptability of messaging content intended to engage suicidal individuals.

Methods: In March 2013, clinic intake staff distributed flyers announcing the study, “Your Feedback Counts” to potential participants (individuals waiting to be seen for a mental health appointment) together with the Patient Health Questionnaire. The flyer explained that a score of two or three (“more than half the days” or “nearly every day” respectively) on the suicide ideation question made them eligible to provide feedback on components of a suicide prevention intervention under development. The patient could access an anonymous online survey by following a link. After providing consent online, participants completed the anonymous survey.

Results: Thirty-four individuals provided data on past demographic information. Participants reported that they would be most drawn to an intervention where they knew that they were cared about, that was personalized, that others like them had found it helpful, and that included examples with real people. Participants preferred email invitations with subject lines expressing concern and availability of extra resources. Participants also provided feedback about a media prototype including a brand design and advertisement video for introducing the intervention.

Conclusions: This paper provides one model (including development of an engagement survey, audience for an engagement survey, methods for presenting results of an engagement survey) for including target users in the development of uptake strategies for online mental health interventions.

(J Med Internet Res 2014;16(2):e42) doi:10.2196/jmir.3173

KEYWORDS

human centered design; user centered design; health 2.0; suicide

Introduction

Internet-based interventions have the potential to increase the accessibility and efficacy of mental health treatment [1-4]. They are effective in producing change; several meta-analyses have found no significant differences in results when compared to face-to-face interventions [1,5,6]. Most of the current, long-established, research-based mental health treatment options serve a population that is already either seeking treatment and receiving care, or is in crisis and getting services from the emergency department [7]. But the Internet helps us broaden the populations being served and reach individuals who might otherwise be missed in more traditional delivery modalities [8]. The Internet can reach those not seeking traditional care due to logistic difficulties in accessing treatment or stigma around mental illness and has potential for large scale dissemination [9]. However, unsatisfactory uptake and follow-through is a significant barrier to reaching the full potential of online interventions [10-13]. Thus, identifying effective strategies to engage the target population is of critical importance for Internet-based mental health interventions to reach their optimum potential.

The involvement of consumer collaborators in mental health research has the potential to transform care and increase patient engagement [14]. Yet, the peripheral role of the end-user in the development of online mental health interventions has been proposed as a major barrier to utilizing such a service or program [15]. Health 2.0 (and Medicine 2.0, ie, the application of Web 2.0 technologies to health care) and human-centered design (a design process extensively relying on the intended user at all levels of the design process) are useful frameworks for enabling users to have an active role in generating and manipulating Web content and participating in health promotion [16,17]. These frameworks provide a theoretical foundation for increasing user engagement and uptake, but deciding how to implement this approach to increase engagement in mental health interventions is challenging.

Suicide is a devastating consequence of mental illness and major mental health concern in the United States, ranking as the 10th leading cause of death and generating high-cost emergency room visits and hospitalizations [18]. Research has led to the production of mental health interventions efficacious in treating suicidality [19-21], including a Dutch online program designed specifically to decrease frequency and intensity of suicidal ideation [22,23]. However, results from the Dutch study indicate that over half the eligible patients did not return the study consent form [22], thus reinforcing the importance of user engagement.

The aim of this project was to take the first step in attempting to increase user uptake and engagement in an online suicide prevention study by soliciting feedback from patients eligible for the intervention program. Specifically, we surveyed patients with current suicidal ideation about acceptable subject lines, intervention descriptions, project names, and introductory videos in order to gain a better understanding of how to reach the target audience.

Methods

Participant Recruitment

Individuals seen for mental health treatment at Group Health, a large health care organization based in Seattle, Washington, are asked to complete the Patient Health Questionnaire (PHQ-9) before every visit as part of routine care. The PHQ-9 is a 9-item, self-report questionnaire used for screening, monitoring, and measuring the severity of depression over the previous 2 weeks [24]. The PHQ-9 uses a 4-point scale (“not at all”, “several days”, “more than half the days”, “nearly every day”), and the last question asks about presence and duration of suicide ideation. An answer of “more than half the days” or “nearly every day” for suicide ideation has been found to predict subsequent suicide attempts or suicide death [25].

In March 2013, clinic intake staff distributed flyers announcing the study, “Your Feedback Counts”, to potential participants (individuals waiting to be seen for mental health appointments) together with the PHQ-9. The flyer explained that a score of two or three (“more than half the days” or “nearly every day” respectively) on the suicide ideation question made them eligible to provide feedback on components of a suicide prevention intervention under development. The patient could access an anonymous online survey by following a link. After providing consent online, participants completed the anonymous survey.

As a last step, participants could link to a separate survey and provide their contact information to receive a US $10 incentive. No link between survey data and contact information was collected. All study procedures were approved by the Group Health Institutional Review Board.

Survey

The survey included questions about demographics and medical treatment for a suicide attempt or self-injury, as well as acceptability of proposed intervention messaging (see Multimedia Appendices 1 and 2). A patient consultant group of 5 individuals with a history of suicidal ideation and suicide attempts helped to generate the messaging options, as well as a video describing the intervention. There were 26 options for message subject lines introducing the intervention, 20 options for intervention characteristics, 9 intervention brand names, and an option for study participants to make additional suggestions. The survey was programmed using DatStat Ilume and accessible by computer or other portable Internet-connected device (smartphone, iPad, etc) [26]. All questions related to the intervention invitation and description were asked using a 5-point Likert scale (0=strongly disagree; 4=strongly agree). In evaluating the results, we calculated the difference in responders endorsing “agree” or “strongly agree” for a particular option and “disagree” or “strongly disagree”. Descriptive statistical analyses were conducted via SPSS 16 [27].

Results

Participants

Of the 39 participants who visited the online consent page, 38 agreed to participate. Three individuals consented but provided no further information and four individuals provided only demographic information. Participants completed the survey in an average of 20.5 minutes (SD 25.98). Most participants were female, had received medical help for a suicide attempt or self-injury during their lifetime, and were under 64 years old (Table 1).

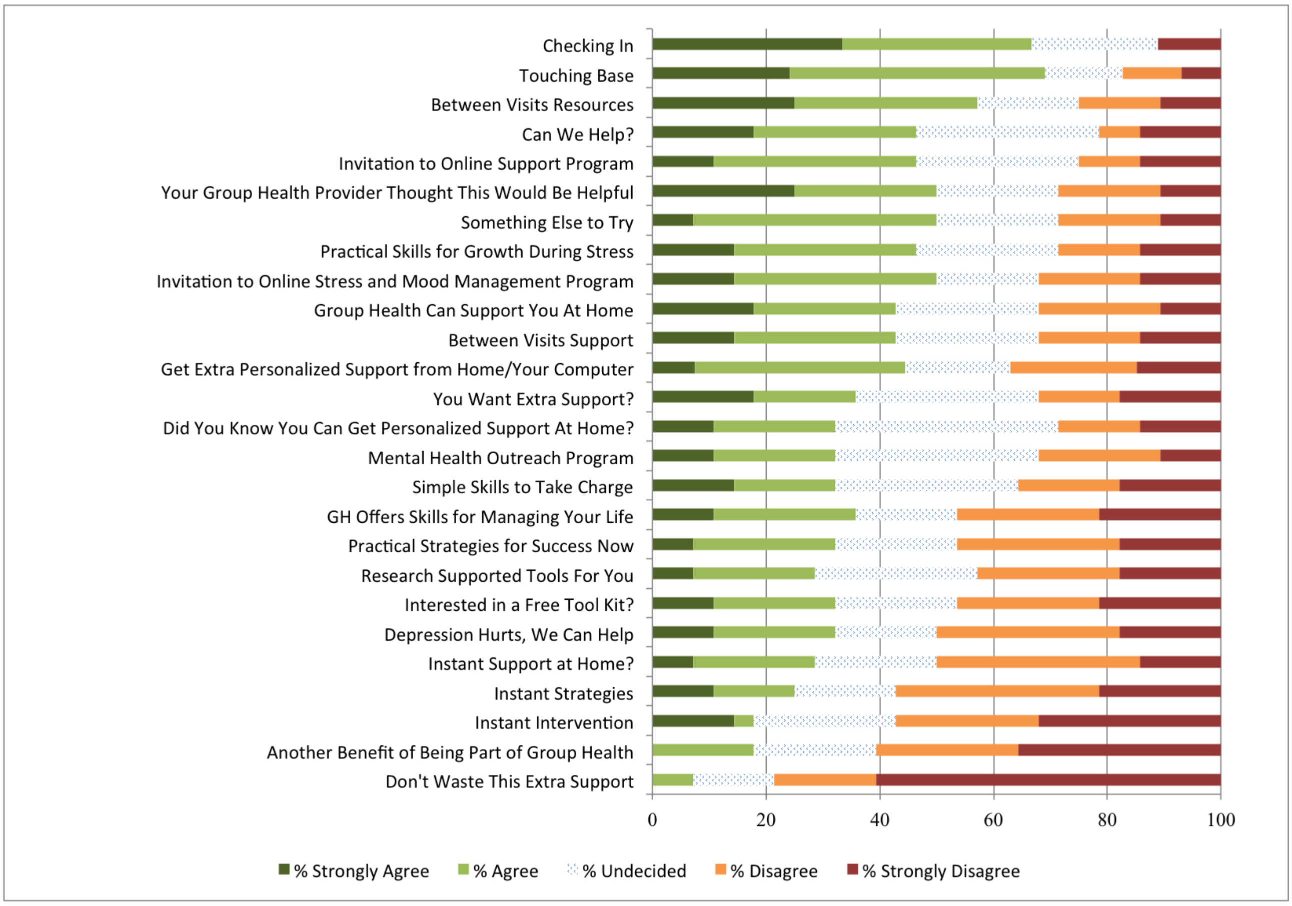
Table 1. Demographics and previous treatment for self-harm (n=34).

Characteristics		n (%)
Age, years		
	<30	16 (47.1)
	30-64	16 (47.1)
	>65	2 (5.9)
Gender		
	Female	23 (67.6)
	Male	11 (32.4)
Receiving medical help for suicide attempt/self-injury (lifetime)		
	Yes	22 (64.7)
	No	12 (35.3)

Invitation Message Subject Line Preferences

Of the 26 different subject lines proposed, 5 emerged as agreeable to more than 45% of responders and disagreeable to less than 25%—“Checking In”, “Touching Base”, “Between Visits Resources”, “Can We Help?”, and “Invitation to Online Support Program”. Messages advertising support as “free” or “quick” (including words as “instant”, “free”, “simple”) were less appealing to participants (Figure 1).

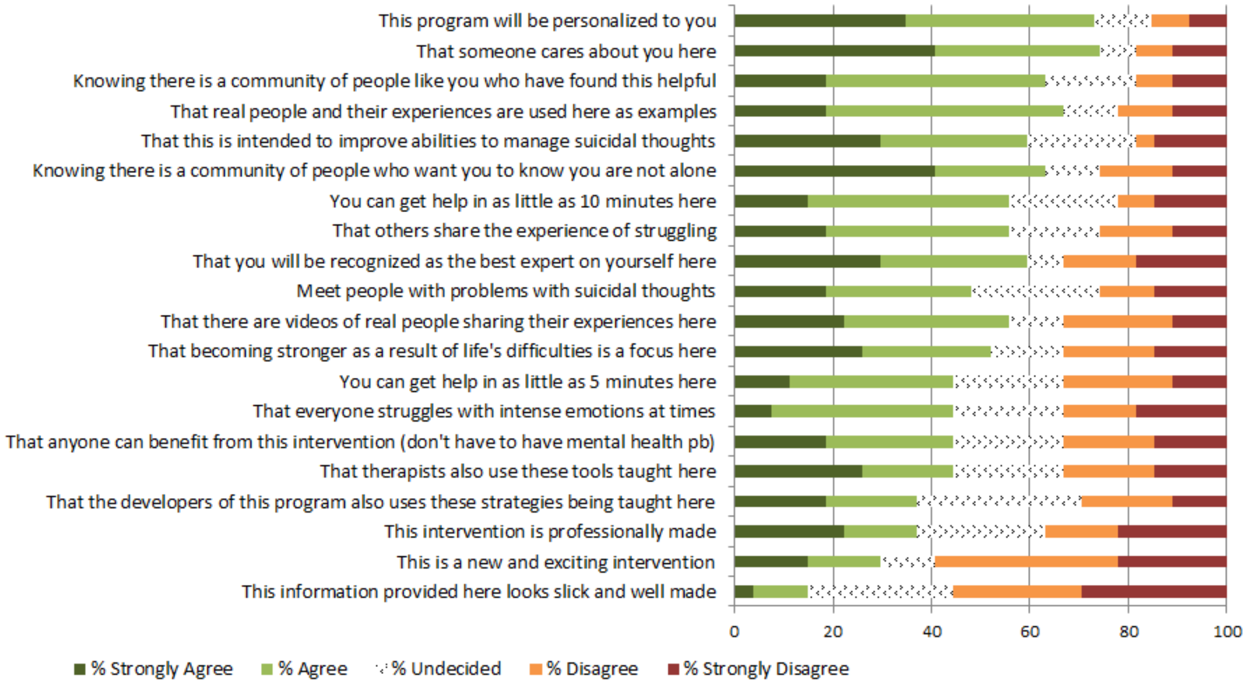
Figure 1. “I would like to open a message with Subject Line...”.



Intervention Characteristic Preferences

Twenty intervention characteristics were provided, and an average of 26.5 participants answered each question (Figure 2). The most preferred characteristics were that the program was personalized, that someone who cares personally is involved, that others have found this program helpful, that real people have found the program helpful, and that "you are not alone". Participants least preferred a "slick" looking and "well made" program.

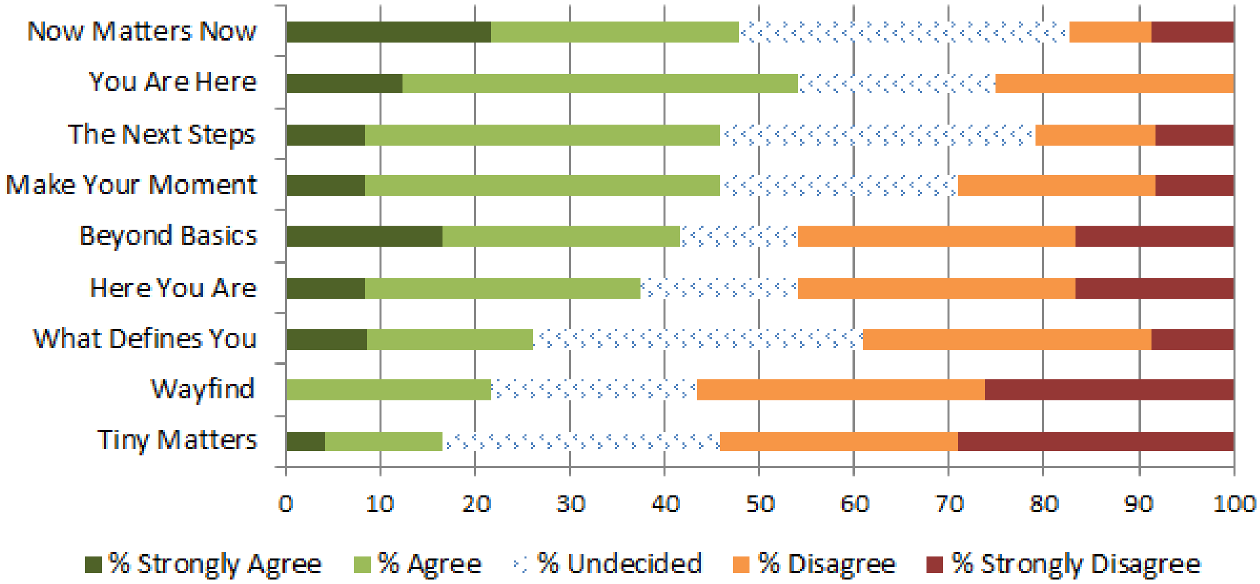
Figure 2. Intervention content preferences.



Intervention Name Brand Preferences

The two most preferred options for the intervention name brand were "Now Matters Now" and "You Are Here". Preference for intervention name varied widely with an agree/disagree gap between –37.5% ("Tiny Matters") and 30.4% ("Now Matters Now"). Each question was answered on average by 23.5 responders. See Figure 3.

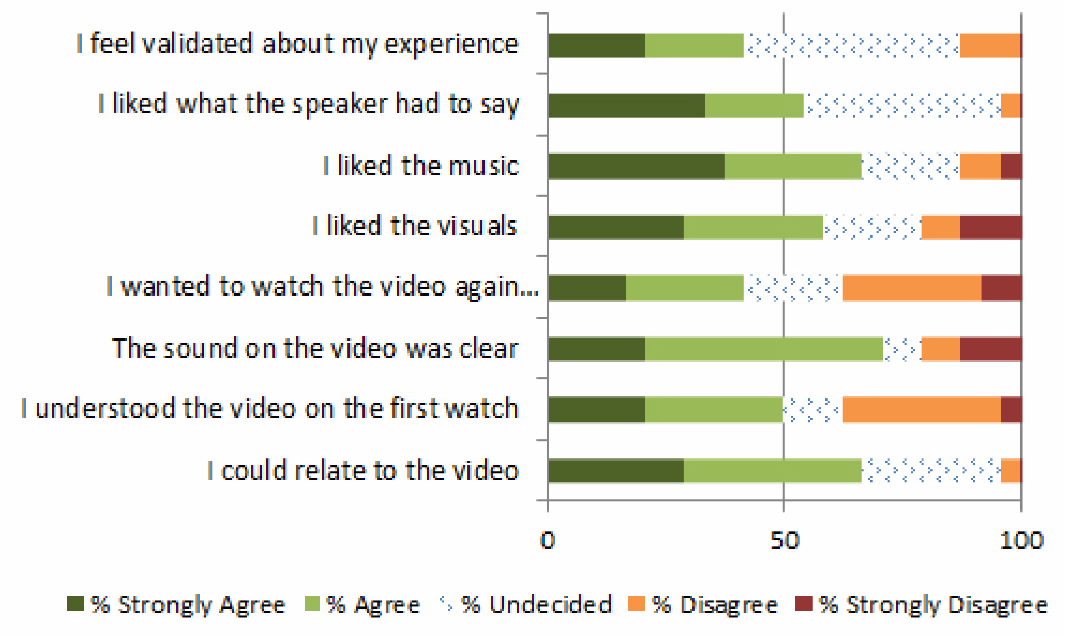
Figure 3. Intervention name brand preferences.



Video Evaluation

Survey participants were asked to watch a 2-minute video about the intervention and answer a series of 8 follow-up questions (see Figure 4). The purpose of the video evaluation was to identify specific aspects of the video that might need improvement. Participants were asked about technical characteristics (eg, quality of sound) and artistic impression (eg, music, visuals) as well as content itself (eg, liking what the speaker had to say).

Figure 4. Feedback on advertisement video.



Discussion

Principal Findings

The main goal of this project was to survey patients with current suicidal ideation about acceptable engagement messaging for an online intervention for individuals with suicidal thoughts. Invitation subject lines, intervention descriptions, project names, and an introductory video were reviewed. The most agreeable subject lines for invitation messages were simple inquiries about how the respondent was doing (“Checking in” and “Touching base”). Also agreeable were messages highlighting additional “resources” that “your provider thought would be helpful”. Certain messaging appeared less agreeable to participants, such as messages that proposed something “free” or “instant”. Intervention characteristics that respondents endorsed were personalization, caring, and real people with personal examples. Less important characteristics were appearance and something “new and exciting”. Two brand names relating to the present moment, “Now Matters Now” and “You Are Here”, were rated most agreeable. Overall, most respondents agreed that they could relate to the video introduction, but responses also indicated room for improvement with technical aspects and video comprehension. This project accomplished the goal of giving researchers a good place to start with messaging designed to promote uptake and engagement in an online intervention. Perhaps, more importantly, we now have a better idea about messaging options to avoid.

Limitations

There are limitations to this project. The sample size was small and demographically narrow and thus did not allow for subgroup analyses or comparisons. However, the study sample demographics were similar to those who most commonly attempt suicide in our health care setting [25]. Individuals in the United States most likely to die by suicide, that is, males in their middle and late years, are not represented here. Future research should apply such branding and messaging questions to this specific population, but also a much larger sample in general, given the large deviations in preferences for intervention name/brand. We chose not to advertise more broadly for this survey (on websites or with advertising, which may have resulted in a larger sample) because we were interested specifically in the population our intervention is targeted to—those at increased risk for suicide attempt in the following year in our health care setting [25].

To preserve the anonymity of responders we asked only basic demographic questions. We do not know how many patients were potentially eligible to participate, so we cannot determine the response rate. Participants were informed of inclusion criteria and then were self-selected into the study, which means there may be potential selection bias. We recruited a treatment-seeking population from a mental-health clinic waiting room, and the target population of the future intervention also includes patients receiving only primary care treatment who may have different messaging preferences. We focused on the brevity of the survey, which limited our ability to explore systematic testing or manipulation of messaging content. An alternative explanation for the findings may be that the messages containing “instant/free/simple” were too vague, rather than the terms themselves being unappealing. Last, the survey options were not based on a theoretical foundation or prior hypotheses due to a lack of prior research targeting suicidal individuals for online interventions.

Future Research

Future research should include systematic manipulation of variations around messaging (eg, comparing “Between visits resources” with “Between visits free resources”) in the hope of providing greater understanding of the effectiveness of the various message components. Future research including larger samples and such manipulations could allow us to determine how interventions might be delivered differently for subgroups of the target population (depending on symptoms, preferences, etc).

Conclusions

We consulted with a group of patients with self-reported suicidal thoughts and past attempts to help generate the options we presented to patients. Patients with mental illnesses and suicidal ideation or behavior often report discrimination on many levels (personal, community, institutional, etc) [28]. Receiving feedback from our patient consults was particularly important since we wanted to avoid messaging that would reinforce the stigma of mental health illnesses and prevent or discourage people affected by mental illness from seeking treatment. This paper provides one model (including development of an engagement survey, audience for an engagement survey, methods for presenting results of an engagement survey) for including target users in the development of uptake strategies for online mental health interventions. Finally, large pragmatic clinical trials should be conducted in order to identify whether an online intervention could reduce suicide attempts or deaths on a large scale.

Acknowledgments

We gratefully acknowledge the funding for this work. Funding was awarded to Dr Whiteside in the form of fellowships from the Group Health Foundation and the American Foundation for Suicide Prevention. This work was also supported by an R-34 from the National Institutes of Mental Health MH097836 awarded to Dr Whiteside.

Conflicts of Interest

None declared.

Multimedia Appendix 1

"Your Feedback Counts"- survey with logos included.

[PDF File (Adobe PDF File), 726KB - jmir_v16i2e42_app1.pdf]

Multimedia Appendix 2

Full survey results.

[PDF File (Adobe PDF File), 357KB - jmir_v16i2e42_app2.pdf]

References

1. Barak A, Hen L, Boniel-Nissim M, Shapira N. A Comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based Psychotherapeutic Interventions. Journal of Technology in Human Services 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
2. Reger MA, Gahm GA. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. J Clin Psychol 2009 Jan;65(1):53-75. [doi: 10.1002/jclp.20536] [Medline: 19051274]
3. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
4. Ybarra ML, Eaton WW. Internet-based mental health interventions. Ment Health Serv Res 2005 Jun;7(2):75-87. [Medline: 15974154]
5. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
6. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
7. Luoma JB, Martin CE, Pearson JL. Contact with mental health and primary care providers before suicide: a review of the evidence. Am J Psychiatry 2002 Jun;159(6):909-916. [Medline: 12042175]
8. Cartreine JA, Ahern DK, Locke SE. A roadmap to computer-based psychotherapy in the United States. Harv Rev Psychiatry 2010;18(2):80-95. [doi: 10.3109/10673221003707702] [Medline: 20235773]
9. Portnoy DB, Scott-Sheldon LA, Johnson BT, Carey MP. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. Prev Med 2008 Jul;47(1):3-16 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.014] [Medline: 18403003]
10. 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]
11. 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]

12. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. J Med Internet Res 2013;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]
13. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
14. Hutchinson A, Lovell A. Participatory action research: moving beyond the mental health 'service user' identity. J Psychiatr Ment Health Nurs 2013 Sep;20(7):641-649. [doi: 10.1111/jpm.12001] [Medline: 23167824]
15. 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]
16. 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]
17. Salvendy G. Handbook of Human Factors and Ergonomics. United States: Wiley; 2012.
18. National Institute of Mental Health. Suicide in the US. URL: http://www.nimh.nih.gov/health/publications/suicide-in-the-us-statistics-and-prevention/index.shtml [accessed 2014-01-10] [WebCite Cache ID 6MWmWlk18]
19. Linehan MM, Comtois KA, Murray AM, Brown MZ, Gallop RJ, Heard HL, et al. Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. Arch Gen Psychiatry 2006 Jul;63(7):757-766. [doi: 10.1001/archpsyc.63.7.757] [Medline: 16818865]
20. Williams JM, Swales M. The use of mindfulness-based approaches for suicidal patients. Arch Suicide Res 2004;8(4):315-329. [doi: 10.1080/13811110490476671] [Medline: 16081399]
21. Townsend E, Hawton K, Altman DG, Arensman E, Gunnell D, Hazell P, et al. The efficacy of problem-solving treatments after deliberate self-harm: meta-analysis of randomized controlled trials with respect to depression, hopelessness and improvement in problems. Psychol Med 2001 Aug;31(6):979-988. [Medline: 11513383]
22. van Spijker BA, Majo MC, Smit F, van Straten A, Kerkhof AJ. Reducing suicidal ideation: cost-effectiveness analysis of a randomized controlled trial of unguided web-based self-help. J Med Internet Res 2012;14(5):e141 [FREE Full text] [doi: 10.2196/jmir.1966] [Medline: 23103835]
23. van Spijker BA, van Straten A, Kerkhof AJ. The effectiveness of a web-based self-help intervention to reduce suicidal thoughts: a randomized controlled trial. Trials 2010;11:25 [FREE Full text] [doi: 10.1186/1745-6215-11-25] [Medline: 20214777]
24. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
25. Simon GE, Rutter CM, Peterson D, Oliver M, Whiteside U, Operskalski B, et al. Does response on the PHQ-9 Depression Questionnaire predict subsequent suicide attempt or suicide death? Psychiatr Serv 2013 Dec 1;64(12):1195-1202. [doi: 10.1176/appi.ps.201200587] [Medline: 24036589]
26. DatStat Illume 5.1.1. 1 ed: DatStat; 2012. URL: http://www.datstat.com/ [accessed 2014-01-28] [WebCite Cache ID 6MyCGZwVs]
27. IBM. SPSS 16.0.: IBM; 2007. URL: http://www-01.ibm.com/software/analytics/spss/ [accessed 2014-01-28] [WebCite Cache ID 6MyCNTSOy]
28. Moskos MA, Olson L, Halbern SR, Gray D. Utah youth suicide study: barriers to mental health treatment for adolescents. Suicide Life Threat Behav 2007 Apr;37(2):179-186. [doi: 10.1521/suli.2007.37.2.179] [Medline: 17521271]

Abbreviations

PHQ: Patient Health Questionnaire

Edited by G Eysenbach; submitted 09.12.13; peer-reviewed by P Quinnett, P Batterham, W van Ballegooijen; comments to author 31.12.13; revised version received 14.01.14; accepted 18.01.14; published 07.02.14.

Please cite as:
Whiteside U, Lungu A, Richards J, Simon GE, Clingan S, Siler J, Snyder L, Ludman E
Designing Messaging to Engage Patients in an Online Suicide Prevention Intervention: Survey Results From Patients With Current Suicidal Ideation
J Med Internet Res 2014;16(2):e42
URL: http://www.jmir.org/2014/2/e42/
doi:10.2196/jmir.3173
PMID:

Original Paper

Healthy Weight Regulation and Eating Disorder Prevention in High School Students: A Universal and Targeted Web-Based Intervention

Megan Jones[1], PsyD; Katherine Taylor Lynch[2], MS, MPH; Andrea E Kass[3], MA; Amanda Burrows[2], MS; Joanne Williams[4], PhD; Denise E Wilfley[3], PhD; C Barr Taylor[1], MD

[1]Laboratory for the Study of Behavioral Medicine, Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, United States

[2]PGSP-Stanford PsyD Consortium, Palo Alto University, Palo Alto, CA, United States

[3]Washington University in St Louis, St Louis, MO, United States

[4]Murdoch Children's Research Institute, Parkville, Australia

Corresponding Author:
Megan Jones, PsyD
Laboratory for the Study of Behavioral Medicine
Psychiatry and Behavioral Sciences
Stanford University School of Medicine
401 Quarry Road, MC 5722
Stanford, CA, 94305-5722
United States
Phone: 1 650 223 5742
Fax: 1 650 724 7389
Email: meganjones@stanford.edu

Abstract

Background: Given the rising rates of obesity in children and adolescents, developing evidence-based weight loss or weight maintenance interventions that can be widely disseminated, well implemented, and are highly scalable is a public health necessity. Such interventions should ensure that adolescents establish healthy weight regulation practices while also reducing eating disorder risk.

Objective: This study describes an online program, StayingFit, which has two tracks for universal and targeted delivery and was designed to enhance healthy living skills, encourage healthy weight regulation, and improve weight/shape concerns among high school adolescents.

Methods: Ninth grade students in two high schools in the San Francisco Bay area and in St Louis were invited to participate. Students who were overweight (body mass index [BMI] >85th percentile) were offered the weight management track of StayingFit; students who were normal weight were offered the healthy habits track. The 12-session program included a monitored discussion group and interactive self-monitoring logs. Measures completed pre- and post-intervention included self-report height and weight, used to calculate BMI percentile for age and sex and standardized BMI (zBMI), Youth Risk Behavior Survey (YRBS) nutrition data, the Weight Concerns Scale, and the Center for Epidemiological Studies Depression Scale.

Results: A total of 336 students provided informed consent and were included in the analyses. The racial breakdown of the sample was as follows: 46.7% (157/336) multiracial/other, 31.0% (104/336) Caucasian, 16.7% (56/336) African American, and 5.7% (19/336) did not specify; 43.5% (146/336) of students identified as Hispanic/Latino. BMI percentile and zBMI significantly decreased among students in the weight management track. BMI percentile and zBMI did not significantly change among students in the healthy habits track, demonstrating that these students maintained their weight. Weight/shape concerns significantly decreased among participants in both tracks who had elevated weight/shape concerns at baseline. Fruit and vegetable consumption increased for both tracks. Physical activity increased among participants in the weight management track, while soda consumption and television time decreased.

Conclusions: Results suggest that an Internet-based, universally delivered, targeted intervention may support healthy weight regulation, improve weight/shape concerns among participants with eating disorders risk, and increase physical activity in high school students. Tailored content and interactive features to encourage behavior change may lead to sustainable improvements in adolescent health.

(J Med Internet Res 2014;16(2):e57) doi:10.2196/jmir.2995

KEYWORDS

healthy weight regulation; universal and targeted delivery; school-based intervention; prevention; adolescents

Introduction

Obesity is a growing problem for children and adolescents in the United States. The 2009-2010 National Health and Nutrition Examination Study data estimated the prevalence of obesity (body mass index [BMI] greater than the 95th percentile for age and sex) as 18.4% for adolescents aged 12-19 years [1]; the rate has increased threefold in the last 30 years [2]. Late adolescence is a key developmental period for establishing lifelong health behaviors, as adolescent overweight can greatly affect health into adulthood [3]. Also, risk for eating disorder onset peaks in adolescence for girls [4], while research suggests that adolescent boys exhibit significant weight-related concerns [5]. Obesity and eating disorders may share risk factors, such as weight and shape concerns, loss of control eating, and unhealthy weight regulation behaviors [6]. Individuals who experience teasing about weight and shape are more likely to develop clinical and subclinical disordered eating symptoms, and adolescents who are overweight report significant weight and shape concerns [6]. These data indicate that, along with the shared risk profile, the overlap between overweight and disordered eating may be substantial and warrant targeted intervention. Finding ways to assess and treat the rise in obesity while sensitively addressing issues related to shape, weight, and body image in adolescents is a key public health goal, and one that has yet to be reached via traditional health programming.

An Internet-based intervention model has unique benefits for obesity and eating disorder prevention, particularly for older adolescents. First, this model allows for universal and targeted intervention delivery. This is critical because by mid to late adolescence, risk factors may be more pronounced, and many adolescents may already be considered overweight or obese or have developed full-syndrome eating disorders. A universal prevention approach that targets cultural norms, policies, and encourages healthy weight regulation behaviors and positive body image among an entire population of students is essential, but may not be sufficient, for students at higher risk or struggling with more serious concerns [7]. Targeted (also referred to as selective) prevention interventions aim to reduce risk factors, in this case, those students already considered overweight who require information and tasks tailored to their unique needs [8]. Note that the term “targeted” as opposed to “indicated” prevention is used because of the notion that weight stabilization in adolescents may be more appropriate prevention strategy in this age group. Targeted interventions for weight gain and eating disorder prevention are also associated with larger effect sizes [9] and thus are a key component of a comprehensive program.

Second, the ability to seamlessly and anonymously provide a universal and targeted intervention can also help mitigate stigma, shame, and teasing, which are common psychosocial problems among adolescents who are overweight [10]. Singling out this group of students *in person* in a school environment is challenging and problematic for a number of reasons, including lack of appropriate resources and increased feelings of stigmatization, which can result in increased probability of dropout.

Third, Internet-based programs are relevant for an adolescent population and may therefore be more acceptable. Adolescence is a time of increased autonomy of food and activity choices and is associated with decreased participation in physical activity. Offering an intervention that complements adolescents’ normal routines (eg, using the computer) may increase adoption and retention. Further, this program does not seek to *increase* the overall time per day that adolescents spend online, but rather, encourage smarter and more effective use of technology to support healthy lifestyle behaviors [11]. Further, as it is delivered during classroom time, it does not add additional sedentary time.

Fourth, delivering interventions online transcends barriers associated with access to care, particularly among certain racial and ethnic minority groups who may not otherwise seek in-person services [12]. Online programs can also greatly enhance the sustainability of the delivery of evidence-based health interventions in school settings, by reducing the costs of providing tailored content for thousands of students at once. Recent research highlights how these programs increase versatility, require relatively low levels of professional support, and can be delivered in a variety of settings [13].

High schools provide an ideal setting in which to provide health promotion programs because of the ability to overcome health care disparities, access issues, and address sociocultural issues (eg, stigma) while meeting educational requirements [14]. However, issues with district policies and program implementation have negatively impacted health program delivery, despite an established need for such a curriculum [15]. Sustainable and cost-effective programs that easily integrate into an existing curriculum may increase adoption and sustainability.

This study describes StayingFit, a 12-session Internet-based program designed to build healthy habits and to support positive body image. Our model uses a universal and targeted approach, by giving youth and families personalized tools designed for use within a community-centered health program. StayingFit uses established principles of behavioral science (eg, self-monitoring and goal setting) to affect sustainable behavior change.

The purpose of this study was to test the feasibility, acceptability, and short-term efficacy of StayingFit for supporting healthy weight regulation and body image improvement among ninth-grade students. Secondary aims were to test for changes in psychosocial variables associated with StayingFit. We hypothesized that StayingFit would produce healthy nutrition (eg, increased consumption of “green” foods), lower levels of sedentary behavior, and reduce weight and shape concerns from baseline to post intervention.

Methods

Study Design

The StayingFit study was an uncontrolled feasibility study using a parallel, nonrandomized design. Twelve high schools were sent letters describing StayingFit and inviting collaboration in the research study. Of these, five schools expressed interest, and due to lack of direct funding for this study, two high schools were selected to participate based on ethnic diversity of the student body. Total enrollment for the San Francisco Bay Area school was 31% white, 4% Asian, 4% African American, 3% Pacific Islander, 0.5% American Indian or Alaska Native, and 2% other; 56% identified as Hispanic/Latino. At the St Louis site, students identified as 71% African American, 28% white, and 1% Asian; 1% of students identified as Hispanic/Latino. Note that for purposes of analysis in this study sample, data were analyzed according to the following racial categories: white/Caucasian, black/African American, multiracial/other, and did not specify. Ethnicity data for students identifying as Hispanic/Latino were also collected.

High school teachers and administration were approached in August 2010 to discuss implementing StayingFit through the physical education course. Enrollment began in February 2011. Approval from the Washington University and Stanford University Institutional Review Boards were obtained.

All ninth-grade students enrolled in physical education classes at the participating high schools during the 2010-2011 and 2011-2012 academic years were offered participation in the research study and completed StayingFit as part of their school’s physical education curriculum. Participation in the study was defined as agreement to undergo baseline and post assessments as indicated by active assent and parental consent by students and parents, respectively.

Measurement periods consisted of baseline and post-intervention assessments, occurring 1 week after the termination of the program. Track assignment was based on BMI percentile data. Students who were overweight (BMI percentile ≥85th for age and sex; Centers for Disease Control [CDC], 2009) were assigned to receive the Weight Management (WM) intervention track, and students who were normal weight (BMI percentile <85th for age and sex; CDC growth curves) were assigned to the Healthy Habits (HH) intervention track. Students were not informed of their track assignment in order to maintain confidentiality, anonymity, and reduce stigma. All students were told that they were participating in StayingFit, a program “for helping adolescents eat well, exercise and maintain a healthy weight, and feel better about their body image”.

Recruitment

Teachers distributed consent forms to students during the first week of each semester, and parents and students were asked to read, sign, and return forms within 1 week. All participants were assigned unique identification numbers.

Intervention

Teachers began delivering the weekly intervention in spring 2011, as part of their physical education course. All students were expected to finish one session per week as part of their class work, with any parts they were unable to complete assigned as homework. Teachers agreed to award class credit for completion of StayingFit sessions, to be determined at their discretion.

StayingFit

The StayingFit program is a 12-session online program promoting healthy weight regulation and improved weight/shape concerns. StayingFit encourages adolescents to take an active role in their personal health attitudes and behaviors. A user-centered design process was used to develop StayingFit, and multiple prototypes, focus groups, and usability testing occurred over the course of several years prior to this study. The core content and structure of StayingFit was adapted from a set of validated programs: Student Bodies [16-18] and Student Bodies-BED [19,20]. Student Bodies is an online eating disorder prevention program for adolescent and college-age women that has been shown to significantly reduce risk, onset, and progression of eating disorders [18]. Student Bodies-BED (binge eating disorders) is a program designed for adolescents at risk for overweight, who may have symptoms of binge eating disorder. It has shown a significant effect on BMI z-score, weight and shape concerns, and binge eating behaviors [20].

The current 12-session program uses a unique screening algorithm to assign youth to individualized programs based on their weight status and eating disorder risk. The use of a validated eating disorder prevention program in the program “core” ensures that a unified message is provided to all students about the importance of developing a positive body image. The core of StayingFit also focuses on nutrition education, which uses the concept of “red” and “green” foods to teach students how to structure their diet [21]. Other key themes woven throughout include incorporating physical activity into daily activities and communicating with peers and family about health. New content in this version of StayingFit adds up-to-date nutrition and physical activity information, interactive online exercises, and modules about weight stigma and social pressures related to body image. See Tables 1 and 2 for a description of the intervention components and weekly themes. To ensure newly learned behaviors are sustained over time and across contexts, StayingFit is linked with family, peer, and school programming. Specifically, the program is delivered in classrooms in conjunction with existing health-oriented curricula, encourages parental and teacher involvement by providing educational newsletters and hosting informational meetings that complement program content, and can be readily integrated within related schoolwide health campaigns.

Table 1. StayingFit intervention components.

Component	Description
Sessions	Each StayingFit session includes 10-15 pages of online content, written at a 9th grade reading level, designed to take approximately 30 minutes to complete.
Learning questions	At the end of each session, students are asked to answer questions about their learning that week. Questions assess knowledge, attitudes, behaviors, and self-efficacy related to specific skills taught in the session. To measure student engagement and enjoyment of the material, students rated the level at which content was helpful, interesting, and fun.
Food log	All students completed a food log, in which students indicated the number of separate servings of fruits, vegetables, other green foods (low fat, high nutrient foods), red foods (high fat, low nutrient foods), and soda they had on the day prior to completing the log. After students submitted the log, they were provided automated feedback related to goals (aligned with USDA guidelines) that they set in the first session of the program.
Meal size log	In the meal size log, students recorded meal times, sizes, and hunger before and after meals. Automated feedback was given based on number of meals and meal size, designed to encourage regular eating.
Physical activity log	Students are asked to record the type, frequency, and intensity of physical activity they completed over the previous week and set physical activity goals for the next week. Automated feedback about their exercise habits is provided.
Weight log (WM Track)	Students in the WM track may also complete a weight log, in which they recorded weekly weight and were provided cautionary feedback if they reported unhealthy weight loss or motivational feedback if they reported weight gain.
Hunger and fullness rating scale (Cohort 2)	A hunger/fullness scale ranging from 0 (starving/ravenous) to 10 (stuffed) was used to teach participants to be more attuned to their internal appetite cues. Participants were encouraged to monitor their hunger level throughout the day and to begin eating when their internal appetite signals reached a hunger level of 3, and to stop eating when they reached a 7.
Discussion board	All students were invited to comment on an anonymous discussion board, accessible 24 hrs a day, but primarily used during class to respond to questions related to program material.
Parent materials	Parents could receive weekly emailed or hardcopy newsletters that provided content coordinated with student sessions.
Teacher materials	Teachers were provided with “StayingFit Program FAQs” and invited to contact the research team with any questions.

Table 2. Weekly themes.

Week	Sessions covered	Topics/content
1	1	Introduce program, describe and educate about “red” and “green” foods, establish individual nutritional intake needs based on Choosemyplate.gov, introduce the concept of appetite monitoring including eating in response to moderate hunger and fullness cues
2	2	Introduce the importance of exercise & how exercise relates to weight management
3	3	Educate about lifestyle activities, set healthy and realistic exercise goals using the FITT principle, create a fitness plan
4	4, 5	4—Introduce the importance of regular eating, educate on reading nutrition labels, encourage cutting out high-calorie drinks 5—Binge eating: what it is and triggers, adding “forbidden foods” into your diet, review the concept of hunger/fullness monitoring, provide a mindful eating exercise
5	6	Body image, self-esteem, direct and indirect triggers of negative thoughts and feelings
6	7	Developing healthy routines with regards to eating, exercise and sleep, overcoming barriers to healthy eating
7	8, 9	8—Making healthy snack choices, serving sizes, why diets do not work and the negative effects of dieting 9—Learn about eating disorders, challenging negative thoughts and cognitive restructuring, feel-good body tips
8	10	Environmental factors that influence eating, eating healthy foods in “risky” situations
9	11, 12	11—Learning how not to participate in stigma about weight, ways to stay confident, strategies for dealing with teasing 12—Maintaining healthy habits in the long run, problem solving

Tracks

StayingFit is divided into two parallel student tracks. The Healthy Habits (HH) track, for students below the 85th percentile of age- and sex-adjusted BMI, describes the goal of the program as developing healthy habits related to nutrition and physical activity. The Weight Management (WM) track, designed for students above the 85th percentile of age- and sex-adjusted BMI, emphasizes eating and exercise for weight maintenance. Track differences are primarily in the language used to describe the content and exercises, rather than the content itself. However, students in the Weight Management track also have access to an optional weight log (see below), if they choose to chart their weight each week.

Sessions

The program pages included photos and interactive exercises designed to increase engagement. These exercises were free-text response questions, pop quizzes, fun facts, or matching games related to session content. Both tracks included interactive self-monitoring logs at the end of each session (see Table 1). Students were asked to complete three logs weekly. All students were invited to comment on an anonymous discussion board set up for StayingFit, on which the research team posted weekly questions related to program material. Students were notified that the discussion board is monitored daily by research team members for questionable content, and any potentially harmful posts are reported to faculty.

Parent Materials

Parent newsletters that provided content coordinated with student sessions were made available in electronic and hardcopy formats.

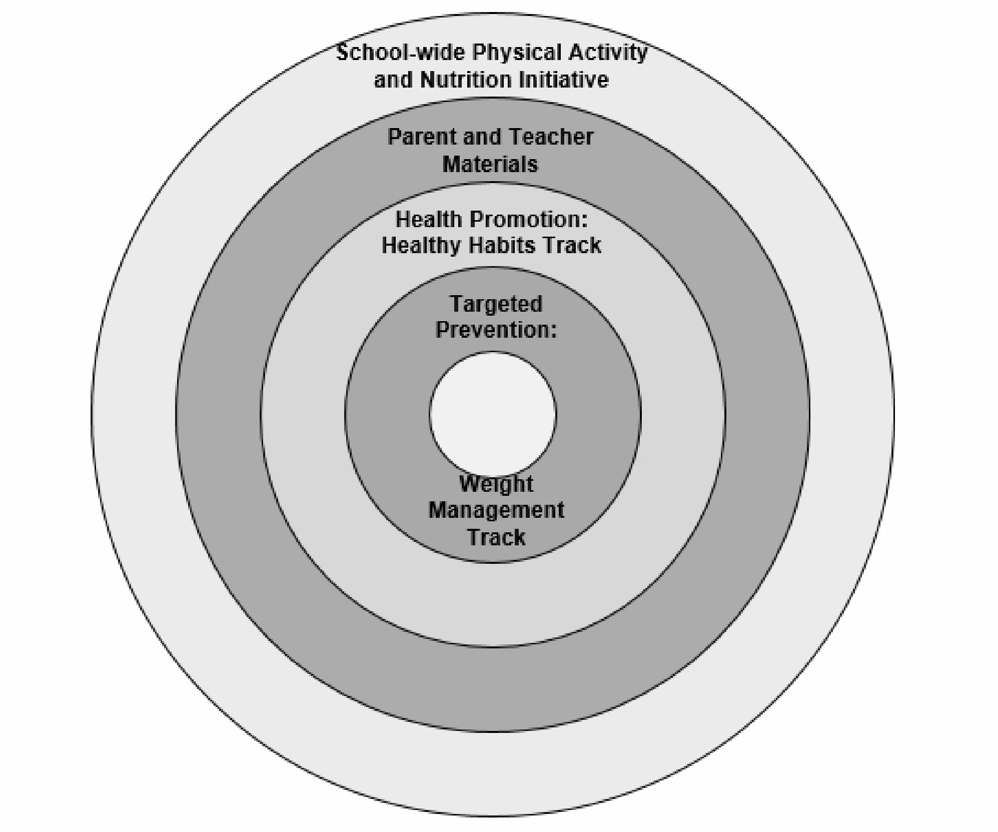
Teacher Materials

In order to reduce required time investment on the part of school staff, StayingFit was designed to be sustainable and easily implementable in classroom settings. To enhance program delivery, teachers were provided with "StayingFit Program FAQs" and invited to contact the research team with any questions.

Schematic

Figure 1 provides a schematic overview of the StayingFit intervention components in the context of the "levels" of intervention. As indicated in the diagram, the program was implemented in the context of school settings that participated in physical activity and nutrition education programs (ie, state and local school district requirements for education on these subjects). The StayingFit intervention directly addressed the levels of parents/teachers and individuals across two intervention levels: no/low risk (health promotion) and high-risk (targeted prevention). Students who required more intensive clinical services (eg, students with anorexia nervosa or with medical problems associated with severe obesity) were recommended to seek outside services.

Figure 1. StayingFit Intervention schematic.



Programming and Privacy

HealthMunk LLC worked with the research team to program StayingFit and provide support throughout research trials. StayingFit is offered on a secured HIPAA-protected secure server and requires login and password credentials to access. Information is encrypted and no identifying information was stored for this study; only students' usernames and passwords were entered into the online interface. Student usernames were chosen by the students, who were instructed not to include any potentially identifying words or phrases in their selection (eg, initials, portions of their name). Passwords were common words assigned by the research team. Students were instructed not to share their usernames or password with peers.

In order to ensure maximum privacy, teachers were not provided with student usernames but were given a list of student passwords, as many students misplaced their original password assignment slips. If students needed to retrieve their usernames, they were instructed to contact the research team.

Measures

The research team visited classrooms at the beginning of each semester to present the program to the students and to help students complete the pre-assessment online via SurveyMonkey. At this time, students and teachers had the opportunity to ask questions. Students were asked to choose individual usernames and passwords to log on to the online interface and given contact information for the research team in case of further questions. The research team set up an email address for centralized communication about the study with parents, teachers, and students. The email was checked daily, and requests for contact were returned within 24 hours. Students completed post-intervention assessments at the end of 12 sessions. All assessments were completed during class time.

Primary Outcomes: Anthropometric Measures

Teachers obtained height and weight 2 weeks prior to the baseline assessment using digital scales and wall measurements of height. At post treatment, students provided self-reported height and weight. Self-report at post treatment was more feasible than arranging for confidential weight and height measurements for all students and was more acceptable to students, teachers, and school administrators. The potential for inaccuracy in self-report anthropomorphic data was adjusted for by using a conservative approach in the analyses, as described in the Statistical Analysis section. BMI percentile and standardized BMI (zBMI) were calculated based on CDC growth charts for age and sex.

Secondary Outcomes: Weight and Shape Concerns

Weight and shape concerns were assessed at baseline and post treatment using the Weight Concerns Scale (WCS) [22]. The WCS assesses worry about weight and shape, fear of gaining 3 pounds (1.36 kg), last time on a diet, importance of weight compared to other life areas, and feelings of fatness. The WCS is significantly correlated with the Eating Disorder Inventory total score (EDI; [23] and the body dissatisfaction subscale of the EDI [24]. The WCS has a test-retest value of approximately 0.85, a 1-year stability of r=.75 [25], and adequate predictive validity [22]. A cut-off score of WCS ≥47 had a sensitivity of 79% and a specificity of 67% for identifying adolescents who developed partial or full EDs [25].

The baseline and post measures also included selected items from the CDC Youth Risk Behavior Survey (YRBS; [26]) to assess nutrition and self-reported physical activity. Clinically meaningful cut-offs were used, based on weight loss recommendations from the Expert Committee, as endorsed by the American Academy of Pediatrics (CDC, 2011). Specifically, these were defined as consuming two or more servings of fruit or vegetables per day; engaging in 5 or more days of physical activity for at least 60 minutes; watching less than 2 hours of television per day; and playing less than 2 hours of video games per day. A cut-off of one or less soda drinks per day was used. The Center for Epidemiologic Studies Depression Scale (CES-D) [27] was used to assess mood.

Statistical Analysis

Analyses were conducted using SPSS version 21.0 (SPSS Inc.). Data were screened for normality. There were large standard deviations in change in BMI percentile and zBMI, due to the use of self-report height and weight data at post intervention. We used a conservative approach of excluding participants with zBMI change greater than 1 standard deviation (SD) above and below the mean. This updated range was compared to the mean zBMI change observed between baseline and post intervention (ie, 0.25 change) in an efficacy trial of the Student Bodies-Binge Eating Disorder online intervention with adolescents [20], on which the StayingFit intervention was based.

Baseline values were imputed for missing post-intervention data (ie, n=71; 21%) on the main outcome variables: BMI percentile and zBMI. Chi-square analyses and t tests were used to examine baseline differences between participants in the two intervention tracks. Paired samples t tests and chi-square tests were used to examine differences from baseline to post test across the outcome variables. Regression analysis was used to examine the effects of intervention track on post-intervention assessment scores, controlling for baseline scores on the same measures. P values less than .05 were considered statistically significant; all tests were two-tailed.

Results

Participants

A total of 336 participants provided assent and their parents provided informed consent. These students were assigned an intervention track based on their BMI percentile. Of these participants, 225 students were assigned to the HH intervention track, and 111 students had been assigned to the WM intervention track. Table 3 provides baseline characteristics for the sample, separated by intervention track. For the total sample, the racial breakdown was as follows: 46.7% (157/336) multiracial/other, 31.0% (104/336) Caucasian, 16.7% (56/336) African American, and 5.7% (19/336) did not specify. 43.5% (146/336) of students also identified as Hispanic or Latino.

Table 3. Participant baseline characteristics, by intervention track.

Variable[a]	HH track (n=225)[b]	WM track (n=111)[b]	*P* value
Age in years, mean (SD)	14.3 (0.63)	14.3 (0.74)	.999
Female, n (%)	137 (60.9)	63 (56.8)	.480
Race, n (%)			
White/Caucasian	91 (40.4)	13 (11.7)	<.001
Black/African American	30 (13.3)	26 (23.4)	.029
Multiracial/Other	94 (41.8)	63 (56.8)	.011
Did not specify	10 (4.4)	9 (8.1)	.210
Ethnicity, n (%)			
Hispanic/Latino	90 (40.0)	56 (50.5)	.079
BMI percentile, mean (SD)	52.8 (22.8)	94.5 (4.12)	<.001
zBMI, mean (SD)	0.04 (0.76)	1.74 (0.43)	<.001
WCS, mean (SD)	26.5 (20.2)	41.4 (19.8)	<.001
CESD, mean (SD)	12.7 (9.0)	15.6 (10.3)	.013
Ate fruit ≥2 times per day, n (%)	48 (24.8)	23 (25.3)	.894
Ate vegetables ≥2 times per day, n (%)	35 (18.1)	16 (17.6)	.370
Drank soda ≥1 time per day, n (%)	29 (15.0)	29 (31.2)	.003
Engaged in physical activity ≥5 days per week, n (%)	112 (58.0)	29 (31.8)	<.001
Watched television ≥2 hrs per day, n (%)	47 (24.3)	44 (48.4)	<.001
Played video games, ≥2 hrs per day, n (%)	47 (24.3)	24 (26.4)	.692

[a] Vegetables=green salad, potatoes, carrots, and other vegetables; Physical activity=at least 60 minutes of physical activity per episode.

[b] n is based on completion of measure at both time points.

Change in Anthropometric Measures

From baseline to post intervention, BMI percentile for participants in the HH intervention track was stable (mean change –0.12 [SD 6.53]; t_{224}=–0.27; *P*=.791), which was an expected finding. However, there was a significant decrease in BMI percentile for participants in the WM intervention track participants (mean change –0.50 [SD 1.49]; t_{110}=–3.51; *P*=.001). There was no significant difference between the intervention tracks on BMI percentile scores at post intervention, controlling for baseline scores on the same measure (standardized beta=.02; *P*=.130).

A similar pattern of results was observed using zBMI. From baseline to post intervention, within-group zBMI change was stable for the HH intervention track participants (mean change –0.007 [SD 0.16]; t_{224}=–0.68; *P*=.497), as expected. However, this change was significant for the WM intervention track participants (mean change –0.03 [SD 0.11]; t_{110}=–2.84; *P*=.005). There was no significant difference between the intervention tracks on zBMI scores at post intervention, controlling for baseline scores on the same measure (standardized beta=–.007; *P*=.565).

Change in Eating and Activity Behaviors

Overview

In total, 193 students (193/225, 85.8%) in the HH track and 91 students (91/111, 82%) in the WM track completed the YRBS measure at both time points.

Consumption of Two or More Servings of Fruit Over the Previous 7 Days

At post intervention, 61 participants (61/225, 31.6%) in the HH tracks and 26 participants (26/111, 28.5%) in the WM track had two or more servings of fruit over the previous 7 days. This was a significant increase from baseline to post intervention for participants in the HH track (χ^2_1=55.7; *P*<.001) and for participants in the WM track (χ^2_1=15.7; *P*<.001).

Consumption of Two or More Servings of Vegetables Over the Previous 7 Days

At post intervention, 41 participants (21.2%) in the HH track and 22 participants (24.1%) in the WM track had two or more servings of vegetables over the previous 7 days. This was a significant increase from baseline to post intervention for participants in the HH track (χ^2_1=15.3; *P*<.001). Change for participants in the WM track was not significant (χ^2_1=0.31; *P*=.752).

Consumption of Soda at Least Once per Day Over the Previous 7 Days

At post intervention, 34 participants (34/225, 17.6%) in the HH tracks and 24 participants (24/111, 26.4%) in the WM track had a soda at least once per day over the previous 7 days. This was a significant increase from baseline to post intervention for participants in the HH track (χ^2_1=13.3; *P*=.001) and a significant decrease for participants in the WM track (χ^2_1=10.5; *P*=.002).

Engaged in 60 Minutes of Physical Activity on at Least 5 of the Previous 7 Days

At post intervention, 85 participants (85/225, 44.0%) in the HH tracks and 31 participants (31/111, 34.1%) in the WM track engaged in at least 60 minutes of physical activity on at least 5 of the previous 7 days. This was a significant decrease from baseline to post intervention for participants in the HH track (χ^2_1=48.4; *P*<.001) and a significant increase for participants in the WM track (χ^2_1=14.9; *P*<.001).

Watched Television for 2 Hours or Less Over Previous 7 Days

At post intervention, 52 participants (52/225, 26.9%) in the HH tracks and 34 participants (34/111, 37.4%) in the WM track watched television for more than 2 hours per day over the previous 7 days. This was a significant increase from baseline to post intervention for participants in the HH track (χ^2_1=38.1; *P*<.001) but a significant decrease for participants in the WM track (χ^2_1=29.7; *P*<.001).

Played Video Games for 2 Hours or Less Over Previous 7 Days

At post intervention, 44 participants (44/225, 22.8%) in the HH tracks and 26 participants (26/111, 28.6%) in the WM track played video games for more than 2 hours per day over the previous 7 days. This was a significant decrease from baseline to post intervention for participants in the HH track (χ^2_1=53.4; *P*<.001) but a significant increase for participants in the WM track (χ^2_1=18.4; *P*<.001).

Change in Psychosocial Variables

Weight and Shape Concerns

From baseline to post intervention, within-group change in weight and shape concerns was not significant for participants in the HH intervention track (mean change –1.55 [SD 12.9]; t_{193}=–1.68; *P*=.095) or for participants in the WM intervention track (mean change –3.02 [SD 15.8]; t_{90}=–1.82; *P*=.071). There was no significant difference between the intervention tracks on weight and shape concerns scores at post intervention, controlling for baseline scores on the same measure (standardized beta=.038; *P*=.327).

We also looked at the subset of participants who had elevated weight and shape concerns (ie, WCS score ≥47) at baseline. There was a significant decrease in weight and shape concerns among participants in the HH track (n=35; mean change –7.23 [SD 17.4]; t_{34}=–2.46; *P*=.019) and among participants in the WM intervention track (n=36; mean change –7.69 [SD 15.4]; t_{35}=–3.00; *P*=.005). However, weight and shape concerns remained above the 47 cut-point at post intervention among both groups.

Depressive Symptoms

From baseline to post intervention, there was a significant within-group increase in depression scores for participants in the HH intervention track (mean change 1.29 [SD 8.14]; t_{167}=2.06; *P*=.041) but there was no significant within-group change for participants in the WM intervention track (mean change 0.67 [SD 10.3]; t_{71}=0.55; *P*=.584). There was no significant difference between the intervention tracks on depression scores at post intervention, controlling for baseline scores on the same measure (standardized beta=.004; *P*=.932).

Discussion

Principal Results

An important objective of this study was to identify potential barriers to implementation and dissemination of a universal and targeted obesity and eating disorders risk reduction program in schools. Acceptability and feasibility of school-based implementation was high. Students and teachers reported satisfaction with program content and implementation and the intervention was inexpensive to deliver. This resulted in the school asking the investigators to offer the program again in subsequent years. Investigators provided assistance with computer problems, but the program was run nearly entirely by the teachers with no training. After the initial cost of program design, content and minor functionality changes could be easily completed.

As this was a pilot study, the primary results are related to short-term behavioral outcomes. BMI significantly decreased among students in the WM track. BMI was stable in the HH track, demonstrating weight maintenance. As has been noted previously, weight maintenance is often an appropriate goal for adolescent obesity prevention as weight stabilization is associated with numerous health benefits [28]. These results are consistent with those found in studies examining previous versions of StayingFit [19,20] and with other research suggesting that Internet-based weight management can be effective for adolescents [29].

The impact of StayingFit on weight and shape concerns was also consistent with previous research demonstrating that the intervention was effective in reducing weight and shape concerns among participants with elevated eating disorders risk. In this case, this finding suggests that StayingFit functioned appropriately as an eating disorder risk reduction intervention while simultaneously supporting health weight regulation. Reducing weight and shape concern among overweight participants is particularly impactful because these individuals are at highest risk for disordered eating behavior.

Eating behaviors also largely improved across tracks. Fruit consumption significantly increased in both groups and vegetable consumption increased significantly for HH participants, but results were not significant for participants in the WM track. The number of participants reporting daily soda consumption significantly increased in the HH group but decreased in the WM group. Similarly split results were observed for other health behaviors. Physical activity significantly decreased among HH participants but increased in WM participants. Increased rates of television viewing were reported by HH participants at post intervention, but WM participants reported a decrease. Interestingly, video game time decreased among the HH group but showed an increase in the WM group. This was not a randomized controlled study, so conclusions and generalizability about the true impact on behaviors are limited. However, the increase in physical activity and decrease in soda consumption and television time among adolescents in the WM track are encouraging. It seems that the StayingFit program was able to have a measurable impact on health habits, in some cases differentially between the two tracks, suggesting that universally delivered interventions can be successful in targeting specific risk behaviors among overweight adolescents.

The potential positive impact on physical activity in the WM track is particularly notable. Teachers, parents, and health practitioners have expressed concern about delivery of health programs online, suggesting that this mode of delivery encourages sedentary behaviors that can contribute to adolescent overweight. However, the research in this area does not support this assertion. In a recent meta-analysis, Internet-based physical activity interventions were shown to have significant effects on exercise behaviors [11]. In fact, these programs have shown promising results in populations identified as sedentary [30], significantly increasing steps per day in a study of overweight adults. These results suggest that online interventions can successfully translate into real-world behavior change. In this implementation, students offered StayingFit also participated in regular physical education classes and were given opportunities to put into practice the physical activity recommendations provided in the program.

However, physical activity decreased among the HH track, suggesting that further support for positive exercise habits is needed. It is possible that students in the HH track were less motivated to change health behaviors due to their normal weight status. Although content in StayingFit aimed to encourage physical activity among all participants, results suggest that this area of the program could be improved possibly capitalizing on competitions and more effective use of social media strategies. In future versions of StayingFit, we plan to include additional apps, games, and personal activity trackers that could further support healthy behavior change among all participants.

The results of this study add further support for the importance of developing Internet-delivered interventions that are dynamic and can be rapidly modified and tailored based on participant monitoring [31]. For example, in this study, participants recorded soda consumption in their self-monitoring logs. Increased soda consumption among the HH group could have triggered a “flag” for add-on supplementary content supporting targeted behavior change. The online format allows for easy intervention modification post hoc. However, to truly maximize the benefits of using technology, these “adjustments” should be made in real time.

Limitations

This study was not intended to be an efficacy trial; hence, no control condition was used nor did randomization or long-term follow-up occur. Outcome data must therefore be interpreted with caution, and regression to the mean is highly possible with this study design. The study also represents implementation at two high schools, and future studies should examine the program feasibility and outcomes across multiple sites.

Future Directions

Adolescent obesity continues to be an issue among US teenagers. Finding comprehensive ways to reduce adolescent overweight through assessment and education about eating, physical activity, and emotional distress may help alleviate health disparities and reduce the rates of common medical and psychiatric comorbidities.

A study utilizing an adaptive design could evaluate further iterative improvements, specifically in how to most effectively make use of in-program process data to tailor interventions [32]. The Internet-based format for health interventions has unique advantages that can be harnessed for positive impact at the individual and populations levels. The “virtual” nature of different program “tracks” allows students in the same classroom to be efficiently “screened” and assigned into intervention versions best suited for their needs. Real-time participant monitoring, meaning “in the moment” observation of program use, is also made more feasible by utilizing program use and self-monitoring journal data to ensure participant safety and provide referral when necessary. Online, moderated discussion groups for participants can leverage the strong influence of social networks to elicit healthy behavior change while still protecting student anonymity. Rather than being a static information source, Internet-delivered interventions harness technology to promote engagement, interactivity, and real-time monitoring and feedback.

While Internet-based approaches may offer a unique and important role in larger obesity and eating disorder prevention efforts, the intervention should not stop at the computer or mobile device. Rather, the technology should be used strategically to overcome limitations of person-based approaches for obesity and eating disorders prevention and facilitate connections with in-person social networks. Technology can aid in connecting individual-level interventions with broader socioecological systems and true universal prevention involving families, schools, communities, and policy change. Future studies should extend the socioecological framework of integrated obesity and eating disorders prevention through partnerships with community organizations and larger school and health systems.

Conclusions

Results suggest that delivery of an Internet-based, universal, and targeted intervention is feasible and effective in a school-based setting. StayingFit appears to support healthy weight regulation, improve weight/shape concerns, and increase healthy food consumption in adolescents.

Acknowledgments

CBT received grant support from the National Institute of Mental Health (NIMH; R01 MH081124) and the National Health and Medical Research Council (NHMRC; 003813). DEW received grant support from NIMH (R01 MH081125 and K24 MH070446). AEK received grant support from the National Heart, Lung, and Blood Institute (NHLBI; T32 HL007456; PI: DEW); JW received grant sponsorship from the National Health and Medical Research Council of Australia Project Grant ID 1003813. The authors thank Healthmunk LLC for providing the software program, and the staff, faculty, and students of the participating high schools for their help with this study.

Conflicts of Interest

None declared.

References

1. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010. JAMA 2012 Feb 1;307(5):483-490. [doi: 10.1001/jama.2012.40] [Medline: 22253364]
2. National Center for Health Statistics (US). Health, United States, 2011: With Special Feature on Socioeconomic Status and Health. Hyattsville, MD: NCHS; 2012.
3. Dietz WH. Health consequences of obesity in youth: childhood predictors of adult disease. Pediatrics 1998 Mar;101(3 Pt 2):518-525. [Medline: 12224658]
4. Striegel-Moore RH, Bulik CM. Risk factors for eating disorders. Am Psychol 2007 Apr;62(3):181-198. [doi: 10.1037/0003-066X.62.3.181] [Medline: 17469897]
5. Kotler LA, Cohen P, Davies M, Pine DS, Walsh BT. Longitudinal relationships between childhood, adolescent, and adult eating disorders. J Am Acad Child Adolesc Psychiatry 2001 Dec;40(12):1434-1440. [doi: 10.1097/00004583-200112000-00014] [Medline: 11765289]
6. Haines J, Kleinman KP, Rifas-Shiman SL, Field AE, Austin SB. Examination of shared risk and protective factors for overweight and disordered eating among adolescents. Arch Pediatr Adolesc Med 2010 Apr;164(4):336-343 [FREE Full text] [doi: 10.1001/archpediatrics.2010.19] [Medline: 20368486]
7. Winzelberg AJ, Eppstein D, Eldredge KL, Wilfley D, Dasmahapatra R, Dev P, et al. Effectiveness of an Internet-based program for reducing risk factors for eating disorders. J Consult Clin Psychol 2000 Apr;68(2):346-350. [Medline: 10780136]
8. Luce KH, Osborne MI, Winzelberg AJ, Das S, Abascal LB, Celio AA, et al. Application of an algorithm-driven protocol to simultaneously provide universal and targeted prevention programs. Int J Eat Disord 2005 Apr;37(3):220-226. [doi: 10.1002/eat.20089] [Medline: 15822091]
9. Stice E, Shaw H, Marti CN. A meta-analytic review of eating disorder prevention programs: encouraging findings. Annu Rev Clin Psychol 2007;3:207-231. [doi: 10.1146/annurev.clinpsy.3.022806.091447] [Medline: 17716054]
10. Neumark-Sztainer D, Martin SL, Story M. School-based programs for obesity prevention: what do adolescents recommend? Am J Health Promot 2000;14(4):232-5, iii. [Medline: 10915534]
11. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
12. Ybarra ML, Eaton WW. Internet-based mental health interventions. Ment Health Serv Res 2005 Jun;7(2):75-87. [Medline: 15974154]
13. Calear AL, Christensen H. Review of internet-based prevention and treatment programs for anxiety and depression in children and adolescents. Med J Aust 2010 Jun 7;192(11 Suppl):S12-S14. [Medline: 20528700]
14. Swinburn BA, Caterson I, Seidell JC, James WP. Diet, nutrition and the prevention of excess weight gain and obesity. Public Health Nutr 2004 Feb;7(1A):123-146. [Medline: 14972057]
15. Baker A. The New York Times. 2012 Jul 11. Despite Obesity Concerns, Gym Classes Are Cut URL: http://www.nytimes.com/2012/07/11/education/even-as-schools-battle-obesity-physical-education-is-sidelined.html?pagewanted=all&_r=0
16. Abascal L, Bruning Brown J, Winzelberg AJ, Dev P, Taylor CB. Combining universal and targeted prevention for school-based eating disorder programs. Int J Eat Disord 2004 Jan;35(1):1-9. [doi: 10.1002/eat.10234] [Medline: 14705151]
17. Bruning Brown J, Winzelberg AJ, Abascal LB, Taylor CB. An evaluation of an Internet-delivered eating disorder prevention program for adolescents and their parents. J Adolesc Health 2004 Oct;35(4):290-296. [doi: 10.1016/j.jadohealth.2003.10.010] [Medline: 15450542]
18. Taylor CB, Bryson S, Luce KH, Cunning D, Doyle AC, Abascal LB, et al. Prevention of eating disorders in at-risk college-age women. Arch Gen Psychiatry 2006 Aug;63(8):881-888 [FREE Full text] [doi: 10.1001/archpsyc.63.8.881] [Medline: 16894064]

19. Doyle AC, Goldschmidt A, Huang C, Winzelberg AJ, Taylor CB, Wilfley DE. Reduction of overweight and eating disorder symptoms via the Internet in adolescents: a randomized controlled trial. J Adolesc Health 2008 Aug;43(2):172-179 [FREE Full text] [doi: 10.1016/j.jadohealth.2008.01.011] [Medline: 18639791]
20. Jones M, Luce KH, Osborne MI, Taylor K, Cunning D, Doyle AC, et al. Randomized, controlled trial of an internet-facilitated intervention for reducing binge eating and overweight in adolescents. Pediatrics 2008 Mar;121(3):453-462 [FREE Full text] [doi: 10.1542/peds.2007-1173] [Medline: 18310192]
21. Epstein LH, Valoski A, Wing RR, McCurley J. Ten-year follow-up of behavioral, family-based treatment for obese children. JAMA 1990 Nov 21;264(19):2519-2523. [Medline: 2232019]
22. Killen JD, Hayward C, Wilson DM, Taylor CB, Hammer LD, Litt I, et al. Factors associated with eating disorder symptoms in a community sample of 6th and 7th grade girls. Int J Eat Disord 1994 May;15(4):357-367. [Medline: 8032350]
23. Garner DM, Olmstead MP, Polivy J. Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. Int J Eat Disord 1983;2(2):15-34. [doi: 10.1002/1098-108X(198321)2:2<15::AID-EAT2260020203>3.0.CO;2-6]
24. Killen JD, Taylor CB, Hayward C, Haydel KF, Wilson DM, Hammer L, et al. Weight concerns influence the development of eating disorders: a 4-year prospective study. J Consult Clin Psychol 1996 Oct;64(5):936-940. [Medline: 8916622]
25. Jacobi C, Abascal L, Taylor CB. Screening for eating disorders and high-risk behavior: caution. Int J Eat Disord 2004 Nov;36(3):280-295. [doi: 10.1002/eat.20048] [Medline: 15478137]
26. Centers for Disease Control. Youth Risk Behavior Survey. 2011. URL: http://www.cdc.gov/yrbs/ [accessed 2014-02-06] [WebCite Cache ID 6NBtBcFub]
27. Radloff LS. The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. J Youth Adolesc 1991 Apr;20(2):149-166. [doi: 10.1007/BF01537606] [Medline: 24265004]
28. Viner R, Bryant-Waugh R, Nicholls D, Christie D. Childhood obesity. Aim should be weight maintenance, not loss. BMJ 2000 May 20;320(7246):1401; author reply 1402-1401; author reply 1403. [Medline: 10858049]
29. Williamson DA, Walden HM, White MA, York-Crowe E, Newton RL, Alfonso A, et al. Two-year internet-based randomized controlled trial for weight loss in African-American girls. Obesity (Silver Spring) 2006 Jul;14(7):1231-1243. [doi: 10.1038/oby.2006.140] [Medline: 16899804]
30. Carr LJ, Bartee RT, Dorozynski C, Broomfield JF, Smith ML, Smith DT. Internet-delivered behavior change program increases physical activity and improves cardiometabolic disease risk factors in sedentary adults: results of a randomized controlled trial. Prev Med 2008 May;46(5):431-438. [doi: 10.1016/j.ypmed.2007.12.005] [Medline: 18207228]
31. 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]
32. Brown CH, Ten Have TR, Jo B, Dagne G, Wyman PA, Muthén B, et al. Adaptive designs for randomized trials in public health. Annu Rev Public Health 2009;30:1-25 [FREE Full text] [doi: 10.1146/annurev.publhealth.031308.100223] [Medline: 19296774]

Abbreviations

BMI: body mass index
CDC: Centers for Disease Control
CES-D: Center for Epidemiologic Studies Depression Scale
EDI: Eating Disorders Inventory
HH: healthy habits
HIPAA: Health Insurance Portability and Accountability Act of 1996
WCS: Weight Concerns Scale
WM: weight management
YRBS: CDC Youth Risk Behavior Survey
zBMI: standardized body mass index

Edited by P Carlbring; submitted 30.09.13; peer-reviewed by R Munoz, L Ritterband, I Cnossen; comments to author 28.10.13; revised version received 22.11.13; accepted 15.01.14; published 27.02.14.

Please cite as:
Jones M, Taylor Lynch K, Kass AE, Burrows A, Williams J, Wilfley DE, Taylor CB
Healthy Weight Regulation and Eating Disorder Prevention in High School Students: A Universal and Targeted Web-Based Intervention
J Med Internet Res 2014;16(2):e57
URL: http://www.jmir.org/2014/2/e57/
doi:10.2196/jmir.2995
PMID:24583683

Original Paper

Motivational Interviewing in a Web-Based Physical Activity Intervention With an Avatar: Randomized Controlled Trial

Stijn Friederichs[1], MSc; Catherine Bolman[1], PhD; Anke Oenema[2], PhD; Janneke Guyaux[1], MSc; Lilian Lechner[1], PhD

[1]Faculty of Psychology and Educational Sciences, Open University of the Netherlands, Heerlen, Netherlands
[2]Department of Health Promotion, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Stijn Friederichs, MSc
Faculty of Psychology and Educational Sciences
Open University of the Netherlands
PO Box 2960
Heerlen, 6401 DL
Netherlands
Phone: 31 0045 5762197
Fax: 31 0045 5762197
Email: stijn.friederichs@ou.nl

Abstract

Background: Developing Web-based physical activity (PA) interventions based on motivational interviewing (MI) could increase the availability and reach of MI techniques for PA promotion. Integrating an avatar in such an intervention could lead to more positive appreciation and higher efficacy of the intervention, compared to an intervention that is purely text-based.

Objective: The present study aims to determine whether a Web-based PA intervention based on MI with an avatar results in more positive appreciation and higher effectiveness of the intervention, when compared to an intervention that is purely text-based.

Methods: A three-arm randomized controlled trial was conducted, containing the following research conditions: (1) a Web-based PA intervention based on MI with an avatar, (2) a content-identical intervention without an avatar, and (3) a control condition that received no intervention. Measurements included PA behavior and process variables, measured at baseline, directly following the intervention and 1 month post intervention.

Results: Both interventions significantly increased self-reported PA at 1 month, compared to the control condition ($\text{beta}_{\text{AVATARvsCONTROL}}$=.39, P=.011; $\text{beta}_{\text{TEXTvsCONTROL}}$=.44, P=.006). No distinctions were found regarding intervention effect on PA between both interventions. Similarly, the results of the process evaluation did not indicate any significant differences between both interventions. Due to the limited relational skills of the avatar in this study, it probably did not succeed in forming a stronger relationship with the user, over and above text alone.

Conclusions: The findings suggest that avatars that do not strengthen the social relationship with the user do not enhance the intervention impact. Future research should determine whether Web-based PA interventions based on MI could benefit from inclusion of a virtual coach capable of more complex relational skills than used in the current study, such as responding in gesture to the user's state and input.

Trial Registration: Dutch Trial Register trial number: NTR3147; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=3147 (Archived by WebCite at http://www.webcitation.org/6NCbwdUJX).

(J Med Internet Res 2014;16(2):e48) doi:10.2196/jmir.2974

KEYWORDS

motivational interviewing; physical activity; Internet; avatar

Introduction

Motivational interviewing (MI) is defined as "a collaborative conversation style for strengthening a person's own motivation and commitment to change" [1]. Evidence illustrates that MI can be successful in getting individuals to increase their physical activity (PA) [2]. Delivering MI in a traditional way, however, is expensive and therefore difficult to scale up. Therefore, developing Web-based PA interventions based on MI could increase the availability and reach of the MI techniques for PA promotion.

Although developing a Web-based PA intervention based on MI appears feasible [3], questions remain to be answered regarding the optimal configuration of such an intervention. MI is usually delivered as a face-to-face intervention by a counselor. In the context of a Web-based delivery mode, the presence of the human counselor could be substituted by using a virtual agent or an avatar [4]. Multiple studies have confirmed that the presence of a virtual agent can further improve effectiveness of Web-based interventions [4,5].

Considering the above, integrating an avatar into a Web-based PA intervention based on MI could lead to more positive appreciation and higher efficacy of the MI components. Future intervention development would be less expensive, however, when such an avatar is omitted. Thus, it is important to determine whether the addition of an avatar leads to more favorable results in the context of a Web-based PA intervention based on MI.

The present study aims to answer the following questions: (1) Does adding an avatar to a Web-based PA intervention based on MI result in additional effects on PA behavior? and (2) Does the presence of such an avatar lead to better appreciation for the intervention?

Methods

Overview

A three-arm randomized controlled trial was conducted, containing the following research conditions: (1) a Web-based PA intervention based on MI with an avatar (AVATAR), (2) a content-identical intervention without an avatar (TEXT), and (3) a control condition that received no intervention (CONTROL). Measurements were taken using Web-based questionnaires at baseline, directly after the intervention (follow-up 1) and 1 month post intervention (follow-up 2).

Participants

The participants were Dutch adults (18-70 years old). Exclusion criteria included impairments that severely impede PA participation (participants were asked whether they were unable to be physically active), not speaking and/or writing Dutch, and not having Internet access. The participants were recruited in April and May 2012 through an Internet panel of Dutch residents who occasionally volunteer in Web-based research.

Measurements

At baseline and follow-up 2, the number of weekly days with at least 30 minutes of moderate PA was measured with a self-reported single item of the Dutch Short Questionnaire to Assess Health Enhancing Physical Activity (SQUASH): "How many days per week are you, in total, moderately physically active, by undertaking, for example, heavy walking, cycling, chores, gardening, sports, or other physical activities for at least 30 minutes?" [6]. At follow-up 1, appreciation for the intervention was evaluated by measuring several appreciation dimensions such as personal relevance, trustworthiness, and overall appreciation.

Interventions

Two Web-based PA interventions based on MI were developed, one of which included an avatar (AVATAR) and one of which was fully text-based (TEXT). Both interventions were derived from a previous study on a Web-based PA intervention based on MI [3]. In these interventions, participants answered several open-ended and multiple choice questions. In between those questions, at various instances, the participants received feedback messages containing a reflection or summary, based on one or more of their prior answers. The following is an example of such a message:

> *So you have a very busy life, and therefore you have less confidence that you could manage to increase your physical activity, which is completely understandable. However, you also said that you're a real go-getter. When you have decided to do something, you go all the way! Because of that, you still have confidence that you could increase your physical activity. Do you already have some ideas about how you would increase your physical activity? What activity would you want to do? And how could you schedule this in such a way that it would not take too much time? With your willpower and a good plan, you would surely be able to become more physically active.*

During both interventions, several topics are discussed such as the participant's current PA behavior, the perceived importance of PA and potential beneficial effects of becoming more physically active, and the participant's confidence that he or she could succeed in becoming more physically active. Finally, participants are given the option to formulate their own PA plans and to anticipate difficulties.

Throughout both interventions, several MI counseling techniques have been implemented and have been translated into automatized text versions. Obviously, application of these techniques by means of an automated Web-based platform differs from application in a real-life counseling setting. For example, a face-to-face setting allows responding to very subtle expressions of motivation, which is less feasible in an automated platform. Due to the specific and interactive approach, however, the application of these techniques in the interventions displays a high degree of similarity to the real-life situation, in which a counselor asks questions and provides feedback. The main difference is that the questions and feedback are not provided by a human counselor, but through text or an avatar on a computer screen.

With regard to textual content, the TEXT and AVATAR interventions are almost completely identical. Throughout both interventions, the feedback messages are identical. During a few moments in the intervention, however, the AVATAR intervention contains somewhat more social dialogue than the TEXT intervention.

The TEXT intervention has a relatively simple layout, consisting of a static blue background and a white dialogue window where questions and feedback messages appear. Figure 1 shows a screenshot of the TEXT intervention.

The AVATAR intervention consists of an avatar positioned behind a desk in a small office. Questions and messages are communicated through text balloons, without the use of voice. The avatar displays speech movements (matching the text in the balloons), social dialogue (at the beginning of the intervention and during transitions between the different parts of the intervention) as well as non-verbal expressions such as empathic gestures and eye and head movements. The physical appearance of the avatar was based on the results of a series of focus group interviews among the target population that discussed how a motivating and reliable avatar should look. Based on the results of these interviews, it was decided to include both a male and a female avatar in the intervention so the participants can choose a coach in accordance with their preferences. At the start of the intervention, participants can choose either a male or a female avatar. Figure 2 shows a screenshot of the AVATAR intervention. Before implementation, the intervention was extensively pretested by members of the target group.

Figure 1. Screenshot of the TEXT intervention.

Voor belangrijkheid koos u het cijfer 5, dus bewegen heeft toch wel enig belang voor u.

Waarom heeft u geen lager cijfer gekozen? Vult u onderstaande zin aan.

Ik vind het toch wel redelijk belangrijk om meer te gaan bewegen, omdat...

Figure 2. Screenshot of the AVATAR intervention.

Voor belangrijkheid koos u het cijfer 5 , dus bewegen heeft toch wel enig belang voor u.

Waarom heeft u geen lager cijfer gekozen? Vult u onderstaande zin aan.

Ik vind het toch wel redelijk belangrijk om meer te gaan bewegen, omdat...

Statistical Analyses

Physical activity data were analyzed using a linear regression analysis with the following independent variables: baseline PA behavior, gender, age, body mass index (BMI), educational level, and the intervention condition variable coded into two dummies (AVATAR and TEXT). Potential variations in process evaluation variables were analyzed through analysis of covariance (ANCOVA) with gender, age, BMI, educational level, and baseline PA behavior as covariates. Analyses were performed using SPSS for Windows (version 18).

Results

At baseline, 958 individuals (60.4% female, 578/958; mean age 42.9 [SD 14.5]; 58% high education, 555/958) completed the questionnaire. These individuals were on average moderately physically active for at least 30 minutes per day an average of 4.2 (SD 1.9) days per week. At baseline, the TEXT condition contained significantly fewer participants with high education compared to the AVATAR and CONTROL condition. No additional baseline differences were found between the three study conditions.

Follow-up 2 (1 month) measurements were completed by 500 participants (AVATAR 162; TEXT 146; CONTROL 192) or 52% of the baseline population (57.8% female, 289/500; mean age 45.3 [SD 14.2]; 58% high education, 290/500). Dropout analyses showed that participants younger than age 46 were more likely to drop out at 1 month (OR 1.95, 95% CI 1.51-2.53). In addition, participants randomized into one of the intervention conditions were more likely to drop out at 1 month (OR 1.64, 95% CI 1.25-2.16).

At 1 month, participants were on average moderately physically active for an average of 4.4 (SD 1.8) days per week for at least 30 minutes per day (AVATAR mean 4.6 [SD 1.6]; TEXT mean 4.7 [SD 1.8]; CONTROL mean 4.0 [SD 1.9]). Both interventions were effective in increasing total PA at 1 month when compared to the control condition ($\text{beta}_{\text{AVATARvsCONTROL}}$=.39, *P*=.011; $\text{beta}_{\text{TEXTvsCONTROL}}$=.44, *P*=.006). Participants from the two intervention conditions who had completed the follow-up 2 questionnaire (n=308) increased the number of days per week on which they were physically active for at least 30 minutes from 4.44 to 4.63 (AVATAR from 4.43 to 4.57; TEXT from 4.46 to 4.69). No differences were found regarding intervention effect on PA level between AVATAR and TEXT condition.

Overall, process evaluation results were quite positive (ie, entertainment 5.16/7; trustworthiness 5.15/7; overall appreciation score 7.14/10). No significant differences were found between the intervention conditions regarding these variables.

Discussion

Principal Findings

Both interventions significantly increased self-reported PA at 1 month, compared to the control condition. These outcomes indicate that Web-based PA interventions based on MI hold promise, as they are potentially capable of inducing behavior change. No distinctions were found regarding effect on PA level between the AVATAR and TEXT intervention. Similarly, the results of the process evaluation did not indicate any significant differences between both interventions.

The avatar in this study did not increase intervention impact. This is probably related to the inability of the avatar to respond in gestures to the user's state and input [5]. Due to the limited relational skills of the avatar in this study, it probably did not succeed in forming a stronger relationship with the user, over and above text alone. As a consequence, the avatar used in the current study was not able to enhance the intervention. For future intervention research, inclusion of an avatar capable of more sophisticated relational skills, such as responding to the user's input with gestures, is warranted.

Alternatively, it would be relevant to test a modality in which the avatar is replaced by multiple short videos of a real human who speaks to the participant and leads the way through the intervention. Previous research shows that this type of interactive video counseling holds promise for public health interventions [7]. These videos could even be supplemented by videos in which a PA expert talks about the possible benefits of being physically active on a regular basis, or in which former participants talk about their experiences during the intervention. All this could help give the Web-based intervention a more human character.

Limitations

This study has some limitations. First, a considerable degree of self-selection occurred in this study, due to the fact that participants from the intervention conditions were more likely to drop out compared to participants from the control conditions. This may be related to a second consideration; participating in this study was relatively demanding since the baseline measurement, intervention, and post-measurement were all on the same day. This may have led to some irritation among the participants and probably explains a part of the relatively large attrition that occurred in the intervention conditions of this study. Third, PA behavior was measured with a single self-report item. Although several studies have provided support for the validity of the item used in this study [8,9], the measure remains a weakness of the current study and an objective measure of PA (eg, by pedometer or accelerometer) is recommended for future studies. Finally, caution is needed when generalizing the results of this study to the general population because of the overrepresentation of a high educational level in the sample due to the Web-based sampling frame [10].

Conclusion

In conclusion, the findings suggest that avatars that do not strengthen the social relationship with the user do not enhance intervention impact. Future research should determine whether Web-based PA interventions based on MI could benefit from inclusion of a virtual coach capable of more complex relational skills than used in the current study, such as responding with gestures to the user's state and input [5]. Furthermore, future research should assess the use of video coaching as a potentially beneficial part of a Web-based PA intervention based on MI.

Acknowledgments

The study was funded by the Netherlands Organization for Health Research and Development (ZonMw, 200120007).

Authors' Contributions

LL, CB, and AO designed and wrote the original proposal. SF, CB, AO, JG, and LL developed the intervention. SF significantly contributed to writing this paper. CB, AO, JG, and LL were involved in revising the manuscript critically. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Miller WR, Rollnick S. Motivational Interviewing. Helping People Change. 3rd edition. New York: The Guilford Press; 2013.
2. Martins RK, McNeil DW. Review of Motivational Interviewing in promoting health behaviors. Clin Psychol Rev 2009 Jun;29(4):283-293. [doi: 10.1016/j.cpr.2009.02.001] [Medline: 19328605]
3. Friederichs SA, Oenema A, Bolman C, Guyaux J, van Keulen HM, Lechner L. Motivational interviewing in a web-based physical activity intervention: questions and reflections. Health Promot Int 2013 Oct 6:-. [doi: 10.1093/heapro/dat069] [Medline: 24101160]
4. Baylor AL. The design of motivational agents and avatars. Education Tech Research Dev 2011 Feb 27;59(2):291-300. [doi: 10.1007/s11423-011-9196-3]
5. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. J Med Internet Res 2012;14(1):e1 [FREE Full text] [doi: 10.2196/jmir.1629] [Medline: 22281837]
6. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003 Dec;56(12):1163-1169. [Medline: 14680666]

7. Jackson RA, Stotland NE, Caughey AB, Gerbert B. Improving diet and exercise in pregnancy with Video Doctor counseling: a randomized trial. Patient Educ Couns 2011 May;83(2):203-209. [doi: 10.1016/j.pec.2010.05.019] [Medline: 21459255]
8. Milton K, Clemes S, Bull F. Can a single question provide an accurate measure of physical activity? Br J Sports Med 2013 Jan;47(1):44-48. [doi: 10.1136/bjsports-2011-090899] [Medline: 22522584]
9. Milton K, Bull FC, Bauman A. Reliability and validity testing of a single-item physical activity measure. Br J Sports Med 2011 Mar;45(3):203-208 [FREE Full text] [doi: 10.1136/bjsm.2009.068395] [Medline: 20484314]
10. Rhodes SD, Bowie DA, Hergenrather KC. Collecting behavioural data using the world wide web: considerations for researchers. J Epidemiol Community Health 2003;57(1):68-73. [doi: 10.1136/jech.57.1.68]

Abbreviations

ANCOVA: analysis of covariance
BMI: body mass index
MI: motivational interviewing
OR: odds ratio
PA: physical activity
SQUASH: Short Questionnaire to Assess Health-Enhancing Physical Activity

---

Edited by G Eysenbach; submitted 24.09.13; peer-reviewed by S Karnes, R Hurling; comments to author 15.11.13; revised version received 26.11.13; accepted 21.01.14; published 13.02.14.

Please cite as:
Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L
Motivational Interviewing in a Web-Based Physical Activity Intervention With an Avatar: Randomized Controlled Trial
J Med Internet Res 2014;16(2):e48
URL: http://www.jmir.org/2014/2/e48/
doi:10.2196/jmir.2974
PMID:24550153

Original Paper

Quit Now? Quit Soon? Quit When You're Ready? Insights About Target Quit Dates for Smoking Cessation From an Online Quit Date Tool

Caroline O Cobb[1], PhD; Raymond S Niaura[1,2,3], PhD; Elisabeth A Donaldson[2], MHS; Amanda L Graham[1,3], PhD

[1]Schroeder Institute for Tobacco Research and Policy Studies, Legacy, Washington, DC, United States

[2]Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[3]Department of Oncology, Georgetown University Medical Center / Cancer Prevention and Control Program, Lombardi Comprehensive Cancer Center, Washington, DC, United States

Corresponding Author:
Amanda L Graham, PhD
Schroeder Institute for Tobacco Research and Policy Studies
Legacy
1724 Massachusetts Ave NW
Washington, DC, 20036
United States
Phone: 1 202 454 5938
Fax: 1 202 454 5785
Email: agraham@legacyforhealth.org

Abstract

Background: Setting a target quit date (TQD) is often an important component in smoking cessation treatment, but ambiguity remains concerning the optimal timing (ie, quitting spontaneously versus delaying to prepare).

Objective: We examined four questions about the timing of TQDs and smoking outcomes in secondary analyses of The iQUITT Study, a randomized trial of Internet and telephone treatment for cessation: (1) What are the characteristics of TQDs set using an online interactive quit date tool?, (2) What are the characteristics of individuals who use a quit date tool and do they differ from those who do not?, (3) Are there differences in smoker characteristics, treatment utilization, and cessation outcomes based TQD timing?, and (4) Is maintenance of an initial TQD predictive of abstinence or do changes to TQDs lead to cessation?

Methods: A total of 825 adult current cigarette smokers were randomized to enhanced Internet or enhanced Internet plus telephone counseling. Latency to TQD in days was calculated as the date difference between the initial TQD and enhanced Internet registration; prospective TQD setters were stratified into four latency groups (0, 1-14, 15-28, 29+ days). Baseline variables, website utilization, and 3-month cessation outcomes were examined between prospective TQD groups. Desire and confidence to quit, number of TQDs, and website logins were tested as predictors of 30-day point prevalence abstinence (ppa) at 3 months (responder-only analyses). Classification and regression tree (CART) analysis explored interactions among baseline variables, website utilization, and latency to TQD as predictors of 30-day ppa.

Results: There were few baseline differences between individuals who used the quit date tool and those who did not. Prospective TQDs were set as follows: registration day was 17.1% (73/427), 1-14 days was 37.7% (161/427), 15-28 days was 18.5% (79/427), and 29+ days was 26.7% (114/427). Participants with a TQD within 2 weeks had higher baseline self-efficacy scores but did not differ on smoking variables. Individuals whose TQD was the same day as registration had the highest logins, page views, number of TQDs set using the tool, and messages sent to other members. Logistic regression revealed a significant interaction between number of TQDs and website logins for 30-day ppa (P=.005). Among those with high logins, 41.8% (33/79) with 1 TQD were abstinent versus 25.9% (35/135) with 2+TQDs. Logins and self-efficacy predicted 30-day ppa in the CART model.

Conclusions: TQD timing did not predict cessation outcomes in standard or exploratory analyses. Self-efficacy and an apparent commitment to an initial TQD were the components most highly related to abstinence but only via interactions with website utilization. Findings highlight the importance of feeling efficacious about handling specific smoking situations and engaging with treatment. Additional research focused on increasing engagement in Web-based cessation studies is needed.

Trial Registration: ClinicalTrials.gov: NCT00282009; http://clinicaltrials.gov/show/NCT00282009 (Archived by WebCite at http://www.webcitation.org/6Kt7NrXDl).

(J Med Internet Res 2014;16(2):e55) doi:10.2196/jmir.3086

KEYWORDS

smoking cessation; Internet; quit date; tobacco dependence

Introduction

Setting a quit date is often a central element of tobacco dependence treatment [1-4]. Establishing a target quit date (TQD) may increase the likelihood of success for several reasons. The public commitment often involved in setting a quit date may increase or solidify a smoker's motivation to quit [5] and the probability that they will follow through with intentions to quit [6]. Setting a TQD may also provide time for the smoker to develop relevant coping skills [6,7] and to obtain and initiate medication use, which can increase the likelihood of abstinence [2,8].

However, there is mixed evidence regarding the importance of the nature (ie, planned vs unplanned) and timing (ie, sooner vs later) of quit dates. Some evidence suggests that setting a TQD is associated with a greater likelihood of making a quit attempt [9] and is a predictor of abstinence [10,11]. Other studies indicate that roughly half of smokers prefer to quit abruptly [12] and do not plan a quit attempt [13-16] and that unplanned or spontaneous quit attempts are more likely to be successful than those involving a TQD [13-17]. It is also unclear whether the timing of a quit date matters. A recent randomized controlled trial by Hughes et al [18] in which smokers were prompted to select a quit date found that those who selected a later quit date or delayed a planned quit attempt were less likely to quit smoking compared to participants who selected an early quit date or adhered to their original date. Similarly, in a trial of varenicline versus placebo for smoking cessation in which smokers chose their own quit dates (within a 5-week time frame), smokers who selected later quit dates (particularly in the last week) were less likely to achieve abstinence in both treatment arms [19]. In contrast, among smokers who planned to quit within a month, proximity of the quit date did not predict abstinence [9]. Similarly, in a Web-based trial by Etter et al [12], smokers randomized to abrupt versus gradual quitting had equivalent quit rates at all follow-ups.

This ambiguity regarding quit dates is reflected in the varying recommendations found on smoking cessation websites. For example, the instructions on the American Cancer Society's website state "Once you've decided to quit, you're ready to pick a quit date. This is a very important step. Pick a day within the next month as your Quit Day" [20]. The American Legacy Foundation's BecomeAnEX website instructs smokers "Don't pick tomorrow as your quit date… Don't set your date too far off in the future… We recommend a day that's about 2-4 weeks away" [21]. The National Cancer Institute's cessation website tells smokers who are preparing to quit to "Pick a date within the next 2 weeks to quit" [22]. QuitAssist, a free website provided by the tobacco company, Altria, simply encourages smokers to "get ready" and "choose a specific quit date" with no specific timeline [23]. For the millions of smokers who search online for assistance quitting smoking [24-26], these mixed messages may be confusing.

Most studies that have examined the timing of a quit date have used retrospective, cross-sectional population-based survey data [13-17] or data gathered in the context of randomized controlled trials in which participants were required to set a quit date or adhere to a researcher-defined date [27,28]. Each of these approaches has limitations. Retrospective reports are subject to recall bias skewed toward remembering more planned quit attempts [29], and required quit dates may not be representative of actual quitting behavior. Prospective research is needed that uses objective methods for measuring the timing of quit dates that occur naturally during the course of quitting.

Web-based cessation programs represent both an effective intervention approach to help smokers quit and a means to address some of the limitations of previous analyses of quitting behavior. Sites that offer interactive tools to assist users in choosing and/or documenting a quit date [30] can yield prospective, naturalistic, and objective measures of quitting behavior with regard to the nature and timing of quit dates. We are aware of only one study that has examined the use of an online quit date tool and its association with abstinence [31].

Our study examined four key questions: (1) What are the characteristics of quit dates that are set using an online interactive quit date tool?, (2) What are the characteristics of individuals who use a quit date tool and do they differ from those who do not?, (3) Are there differences in smoker characteristics, treatment utilization, and cessation outcomes based on the timing of an initial (ie, first) TQD in relation to program initiation?, and (4) Is the maintenance of a TQD predictive of eventual abstinence, or are multiple changes of an online quit date more likely to lead to cessation? We approached these questions in secondary analyses of data from a pragmatic randomized trial of Internet and telephone treatment for smoking cessation [32]. Participants were not required to set a quit date and could use the website as they desired. We began with standard analytic methods to describe differences among those who used an online interactive quit date tool and those who did not. We then examined differences among prospective quit date setters based upon the latency to an initial TQD. We hypothesized that individuals whose target quit date was within 2 weeks of registration would be more motivated to quit, have higher indices of treatment utilization, and be more likely to maintain abstinence. We also hypothesized an interaction between the number of TQDs set and website utilization, such that the highest abstinence rates would be observed among participants with only one TQD (signaling unwavering commitment) and high levels of website utilization. To guide future studies, we employed an exploratory data analysis technique, classification and regression tree analysis (CART) [33], to examine the interactive nature of various predictors on abstinence. This exploratory method can augment traditional analytic approaches to identify unique combinations of variables related to tobacco use behavior patterns [34,35].

Methods

Participants

Participants in The iQUITT Study [32,36] were smokers aged 18 and older in the United States who smoked 5 or more cigarettes per day. To maximize generalizability of study findings, motivation to quit and willingness to set a quit date were not included as eligibility criteria. Active user interception sampling was used to recruit smokers who used the terms “quit(ting) smoking”, “stop(ping) smoking”, or “smoking” in a major Internet search engine and who clicked on a link to QuitNet, the cessation treatment website being evaluated [37]. Following online informed consent and a baseline telephone assessment, participants were randomized to basic Internet, enhanced Internet, or enhanced Internet plus telephone counseling in the parent trial. Follow-up assessments were conducted by phone or online for telephone non-responders at 3, 6, 12, and 18 months. These analyses focus on participants with complete 3-month follow-up data in the two enhanced Internet arms, which included an interactive tool to assist users in setting a quit date (“Quit Date Wizard”). The basic Internet intervention did not include the Quit Date Wizard. Across both enhanced Internet arms, 75% (995/1326) of participants were reached at 3 months. Due to a technical issue early in the trial, data on use of the Quit Date Wizard were not stored for 170 participants. Thus, the final sample for these analyses focused on 825 participants (412 enhanced Internet, 413 enhanced Internet plus telephone counseling).

Interventions

Participants randomized to enhanced Internet were given 6 months of free access to the premium service of the QuitNet website. QuitNet is a widely used Internet cessation program that incorporates evidence-based elements of tobacco dependence treatment [2] including practical counseling and tailored information for cessation, recommendations and support for approved pharmacotherapy, and intra-treatment social support through a large online social network [36,38,39].

The Quit Date Wizard is a central feature of QuitNet. It explains the importance of setting a quit date and prompts users to think about a realistic time frame for quitting (“To choose a timeframe, think about approximately when you will be ready to quit”) with options ranging from “In a week” to “In more than 2 months”. The Wizard also encourages users to consider potential triggers, steps to prepare to quit, and pharmacotherapy use. The Quit Date Wizard does not specify an optimal timeframe for quitting but encourages users to consider whether they feel prepared and if not “to spend a few weeks getting to the point where you are comfortable with the idea of ‘jumping in’ [to quitting]”. Users can enter their own date or select a Wizard-generated quit date. Users can also make their quit date visible to other members for support and can sign up for quit support emails timed around their quit date. Repeated reminders to set a quit date using the Quit Date Wizard or to confirm a previously set quit date are featured prominently throughout QuitNet. Users can update their quit date at any time. These analyses focus on the initial TQD, measured as the number of days between website registration and the first TQD that the user set in the Quit Date Wizard. We elected to examine this TQD versus subsequent updates or changes to a quit date to inform recommendations provided by Internet smoking cessation programs. These analyses are not designed to address the timing of a quit date subsequent to a slip or relapse.

Participants randomized to enhanced Internet plus telephone counseling were offered 5 calls in a relapse-sensitive schedule [40]. Counselors had real-time access to summary data regarding a participant’s use of the QuitNet site, which enabled them to prompt and reinforce use of QuitNet (including the Quit Date Wizard) during each call.

Data Collection and Measures

Summary

The three sources of data are described below. These analyses focus on 3-month data since study questions addressed initial quitting behavior, and this is typically where treatment utilization and intervention effects are the strongest.

Baseline Assessment

Age, gender, race, ethnicity, education, employment, and household income were assessed. We also assessed self-rated health status [41], history of smoking-related illness, body mass index, and whether they had spoken to a doctor about their smoking. Smoking variables included cigarettes per day, the time to first cigarette item from the Fagerström Test for Nicotine Dependence [42], duration of last quit attempt (days), desire to quit and confidence in quitting (scale=1-10), spouse smoking status, and number of smokers in the home. Psychosocial items included the Smoking Situations Confidence Inventory and the Smoking Temptations Inventory (short-form) [43] as measures of self-efficacy, the Perceived Stress Scale [44], the Center for Epidemiologic Studies-Depression (CES-D) Scale [45], Weight Concern Scale [46], the Social Network Index [47], an abbreviated version of the Partner Interaction Questionnaire [48,49], and an item from the Two-Item Conjoint Screen [50] assessing alcohol consumption.

Three-Month Follow-Up Assessment

Smoking outcomes included number of intentional quit attempts and 30-day point prevalence abstinence (ppa; primary outcome of the parent trial) calculated using responder-only analyses. Participants also reported use of other quit methods since enrolling in the trial, including nicotine replacement therapy, behavioral treatment (eg, self-help materials, individual counseling), and prescription medication use (eg, bupropion).

Treatment Utilization

Website utilization metrics included date of QuitNet registration, date of initial TQD, total number of quit dates set using the Quit Date Wizard, website logins, page views, total time online, exchange of messages with other QuitNet members (yes/no), and use of an interactive Medication Wizard (yes/no). Number of calls completed was examined among individuals randomized to enhanced Internet plus telephone counseling.

Statistical Analyses

For Study Question 1, frequency counts were used to characterize use of the Quit Date Wizard. Latency to TQD (days) was calculated as the difference between the first date designated using the Quit Date Wizard and the website registration date. To anchor our analyses to common recommendations provided to smokers in Web-based cessation programs, we categorized this variable as 0 days (registration day), 1-14 days (within 2 weeks), 15-28 days (2-4 weeks), and 29+ days (more than 4 weeks). For Study Question 2, selected baseline characteristics of QuitNet registrants were compared between those who set a quit date using the Quit Date Wizard and those who never set a quit date. For Study Question 3, selected baseline characteristics, treatment utilization metrics, and smoking outcomes were examined by latency to TQD using the categories described above: 0 days, 1-14 days, 15-28 days, and 29+ days. We report the median and interquartile range for skewed variables. Between-group comparisons of categorical items and skewed variables were analyzed using nonparametric statistics, and continuous items were analyzed with analysis of variance (ANOVA) using IBM SPSS (version 21.0). For Study Question 4, a logistic regression model examined 30-day ppa as the primary outcome, number of quit dates set using the Quit Date Wizard, number of logins, and the interaction term (centered at the mean) as predictors, and treatment group, desire to quit, and confidence in quitting as covariates using JMP (version 10.02). We examined Study Questions 1-3 by treatment arm and found no between group differences on likelihood of use of the Quit Date Wizard, latency to TQD, baseline characteristics, or website utilization metrics. Therefore, we combined participants from both treatment arms and report the results for the combined sample.

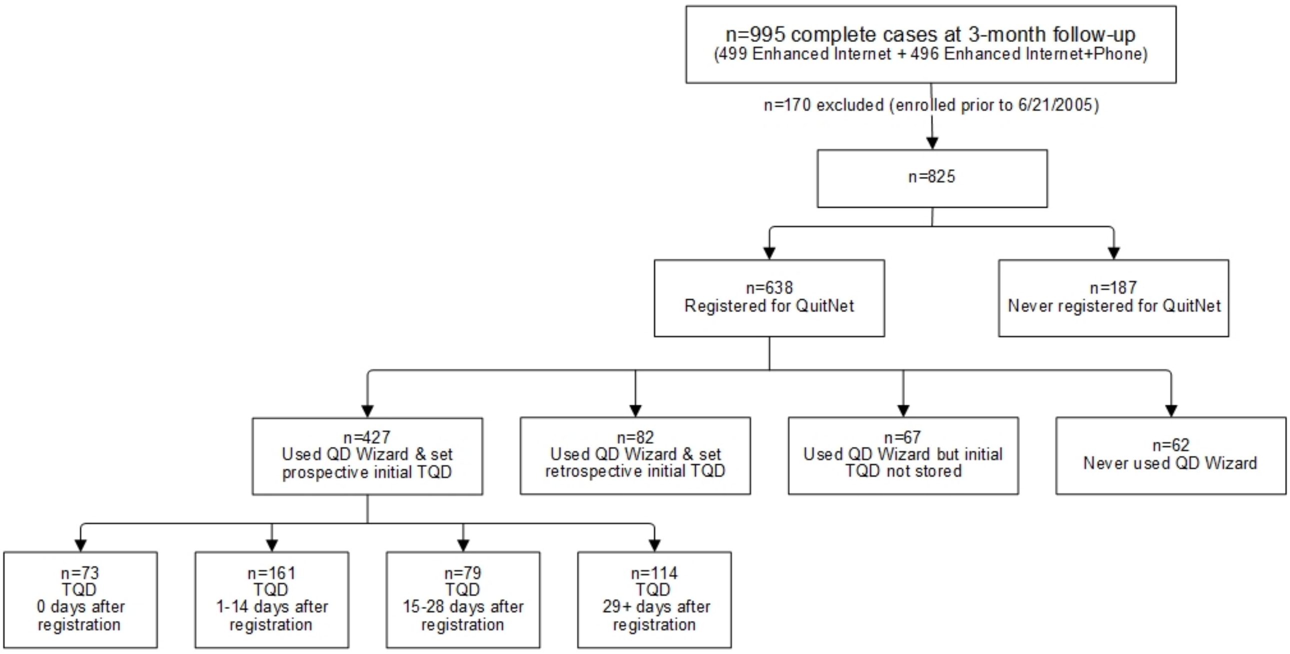
Classification and regression trees (CART) analysis was performed in JMP (version 10.02) to explore the effects of study condition, all baseline variables, and selected treatment utilization measures (logins, number of quit dates set using the Quit Date Wizard, Medication Wizard use, latency to TQD, behavioral treatment use, and pharmacotherapy use) on 30-day ppa, the main outcome of the parent trial [32]. CART analysis allows for a flexible format in terms of allowable response and predictor variables, and handling of missing data [33]. CART is a machine-learning approach that utilizes a classification algorithm to split data into binary subgroups (branches) based upon predictor variables in order to maximize the homogeneity of the two samples for the outcome of interest. In JMP, binary splits for a categorical dependent variable (Y) like abstinence (yes, no) are determined by maximizing the LogWorth statistic ((-log10(*P* value)) [51]. The factors (X; predictors) can be either continuous or categorical (nominal or ordinal). If X is continuous, then the partition is done according to a splitting "cut" value for X. If X is categorical, then it divides the X categories into two groups of levels and considers all possible groupings in two levels. Our CART model included all predictor variables entered simultaneously. To gauge the reliability of our CART analyses, we utilized a k-fold cross-validation procedure that divides the data into k subsets (in this case k=5) that are used to validate the model fit on the rest of the data, fitting a total of K models. The model giving the best validation statistic (-2LogLikelihood) is chosen as the final model.

Results

Study Question 1

Among all participants, 77.3% (638/825) registered on QuitNet following randomization and 22.7% (187/825) never registered (Figure 1). Among QuitNet registrants, 66.9% (427/638) used the Quit Date Wizard to prospectively set a TQD, 12.9% (82/638) used it to record a TQD that occurred prior to registration (retrospective), and 9.7% (62/638) did not use the tool at all. For 10.5% (67/638) of registered participants, use of the Quit Date Wizard was documented but TQDs were not stored due to a database error.

Figure 1. Diagram of data flow from complete cases randomized to the enhanced Internet and enhanced Internet plus telephone counseling arms based on QuitNet registration status, Quit Date (QD) Wizard use, and initial target quit date (TQD) status, and latency to TQD relative to registration date (among prospective quit date setters).



Study Question 2

Compared to those who used the Quit Date Wizard (n=576), those who did not (n=62) were more likely to be male (67.7%, 42/62 vs 49.3%, 284/576, *P*=.009) and to have a spouse who smokes (64.9%, 24/37 vs 45.5%, 157/345, *P*=.039). There were no differences on smoking variables, including smoking rate, desire to quit, or confidence in quitting (Multimedia Appendix 1).

Study Question 3

Among those who set a prospective TQD (n=427), 17.1% (73/427) picked the same day as registration, 37.7% (161/427) picked a date 1-14 days later, 18.5% (79/427) picked a date 15-28 days later, and 26.7% (114/427) picked a date 29 or more days later (see Figure 1). There were differences between prospective TQD groups on education (*P*=.040) and the Confidence Inventory (*P*=.045) (Table 1). There were also differences between groups on treatment utilization metrics. Individuals whose TQD was the same day as registration had the highest number of logins, viewed more webpages, and set more TQDs using the Quit Date Wizard relative to other groups. This group was also the most likely to use one-to-one messaging (31.5%, 23/73) and the least likely to use the Medication Wizard (16.4%, 12/73). Among those who reported at least one quit attempt at the 3-month follow-up (total 389/425; 2 missing cases), there were no differences in use of behavioral quit methods, pharmacotherapy, or telephone counseling calls completed based upon latency to TQD. There were no differences based on latency to TQD on cessation outcomes (Table 2). Overall, 30-day ppa was 21.1% (90/426), and 91.5% (389/425) reported at least one quit attempt.

Table 1. Baseline characteristics by latency to target quit date (TQD) relative to website registration date.

Baseline variable	TQD, 0 days n=73	TQD, 1-14 days n=161	TQD, 15-28 days n=79	TQD, 29+ days n=114	*P* value[a]
Demographic variables [b]					
Age, (years), mean (SD)	34.22 (10.37)	36.99 (10.97)	36.90 (9.98)	38.61 (11.49)	.064
Gender (Female), n (%)	36 (49.3)	84 (52.2)	37 (46.8)	53 (46.5)	
Race, n (%)					.328
White	66 (90.4)	145 (90.1)	70 (88.6)	95 (83.3)	
Non-white	7 (9.6)	16 (9.9)	9 (11.4)	19 (16.7)	
Ethnicity (Hispanic), n (%)	1 (1.4)	8 (5.0)	3 (3.8)	2 (1.8)	.363
Education, n (%)					.040
High school or less	9 (12.3)	33 (20.5)	22 (27.8)	37 (32.5)	
Some college	42 (57.5)	76 (47.2)	34 (43.0)	43 (37.7)	
College 4+ yrs	22 (30.1)	52 (32.3)	23 (29.1)	34 (29.8)	
Employment, n (%)					.616
Employed fulltime	50 (68.5)	114 (70.8)	55 (69.6)	87 (76.3)	
Other[c]	23 (31.5)	47 (29.2)	24 (30.4)	27 (23.7)	
Income, n (%)					.409
Low income (≤$40,000)	33 (45.2)	69 (42.9)	36 (47.4)	60 (53.1)	
High income (>$40,000)	40 (54.8)	92 (57.1)	40 (52.6)	53 (46.9)	
Smoking variables					
Cigarettes per day, mean (SD)	20.26 (10.15)	18.75 (7.90)	20.80 (9.32)	19.61 (9.32)	.356
Time to first cigarette, n (%)					.825
Within 30 minutes	57 (78.1)	122 (75.8)	57 (72.2)	88 (77.2)	
After 30 minutes	16 (21.9)	39 (24.2)	22 (27.8)	26 (22.8)	
Duration of last quit attempt, n (%)					.497
≤3 days	34 (49.3)	82 (55.0)	44 (59.5)	59 (60.2)	
4+ days	35 (50.7)	67 (45.0)	30 (40.5)	39 (39.8)	
Desire to quit, mean (SD)	9.25 (1.08)	9.07 (1.24)	8.87 (1.25)	8.95 (1.43)	.263
Confidence in quitting, mean (SD)	6.49 (2.09)	6.48 (2.13)	5.72 (2.28)	6.16 (2.08)	.052
Psychosocial variables					
Health status, n (%)					.558
Excellent	12 (16.4)	13 (8.1)	7 (8.9)	11 (9.7)	
Very good	26 (35.6)	63 (39.1)	26 (32.9)	40 (35.4)	
Good	20 (27.4)	57 (35.4)	31 (39.2)	35 (31.0)	
Fair/Poor[d]	15 (20.5)	28 (17.4)	15 (19.0)	27 (23.9)	
Illness caused by smoking, n (%)	45 (61.6)	105 (65.2)	40 (51.3)	59 (51.8)	.069
Spouse smokes, n (%)	20 (51.3)	40 (42.6)	23 (45.1)	31 (48.4)	.787
1+ smokers in house, n (%)	17 (23.3)	23 (14.3)	11 (13.9)	27 (23.7)	.103
Temptations Inventory, mean (SD)	4.00 (0.47)	3.90 (0.49)	3.90 (0.59)	3.92 (0.52)	.567
Confidence Inventory, mean (SD)	2.88 (0.60)	2.82 (0.57)	2.67 (0.48)	2.71 (0.57)	.045
Perceived Stress Scale, mean (SD)	6.10 (3.11)	5.90 (2.91)	6.25 (3.26)	6.89 (3.18)	.067
CES-D Scale, mean (SD)	8.79 (5.78)	8.73 (5.27)	9.96 (6.12)	10.31 (5.81)	.082

[a] Nonparametric test (categorical) or ANOVA used.

[b] Participants could refuse to answer a question or respond “I don’t know”. Income, n=423; duration of last quit attempt, n=390; health status, n=426; illness caused by smoking, n=426; spouse smokes, n=248 (asked only among individuals with spouse).

[c] Includes part-time employment, retired, student, homemaker, and unemployed.

[d] Collapsed “Fair” and “Poor” categories due to small cell counts.

Table 2. Treatment utilization and smoking outcomes at 3 months by latency to target quit date (TQD) relative to website registration date.

Dependent measure	TQD, 0 days n=73	TQD, 1-14 days n=161	TQD, 15-28 days n=79	TQD, 29+ days n=114	*P* value[a]
Website utilization					
Logins, n (%)					.024
1-2	14 (19.2)	51 (31.7)	17 (21.5)	37 (32.5)	
3-5	11 (15.1)	29 (18.0)	26 (32.9)	28 (24.6)	
6-10	18 (24.7)	28 (17.4)	9 (11.4)	19 (16.7)	
More than 10	30 (41.1)	53 (32.9)	27 (34.2)	30 (26.3)	
Page views, median (interquartile range)	138 (362)	102 (198)	98 (256)	59.50 (158)	.016
Total number minutes online, median (interquartile range)	88 (237)	62 (157)	54 (150)	43 (119)	.212
Number of quit dates set using Quit Date Wizard, mean (SD)	2.44 (1.73)	1.95 (1.42)	1.72 (0.97)	1.57 (1.40)	.002
Used one-to-one messaging, n (%)	23 (31.5)	38 (23.6)	16 (20.3)	15 (13.2)	.023
Used Medication Wizard, n (%)	12 (16.4)	53 (32.9)	30 (38.0)	35 (30.7)	.025
Other treatment utilization at 3 months (among those who made a quit attempt, n=389) [b]					
Used pharmacotherapy, n (%)[c]	38 (55.9)	93 (63.7)	47 (61.8)	52 (53.1)	.350
Used behavioral treatment, n (%)[d]	10 (14.7)	33 (22.6)	23 (30.3)	21 (21.4)	.167
No. counseling calls completed, mean (SD)[e]	3.39 (2.72)	4.43 (2.87)	4.47 (2.58)	4.83 (3.03)	.159
Smoking outcomes [f]					
30-day ppa, n (%)	19 (26.0)	33 (20.6)	18 (22.8)	20 (17.5)	.555
No. quit attempts, n (%)					.158
0	4 (5.5)	14 (8.8)	3 (3.8)	15 (13.3)	
1	22 (30.1)	50 (31.3)	29 (36.7)	30 (26.5)	
2	24 (32.9)	33 (20.6)	19 (24.1)	21 (18.6)	
3	8 (11.0)	26 (16.3)	10 (12.7)	25 (22.1)	
4+	15 (20.5)	37 (23.1)	18 (22.8)	22 (19.5)	

[a] Nonparametric test (median; categorical) or ANOVA used.

[b] Participants could refuse to answer a question or respond “I don’t know”. Pharmacotherapy, n=388; used behavioral treatment, n=388.

[c] NRT, Zyban, Chantix.

[d] Individual counseling, group counseling, pamphlet/books, telephone counseling not through the study.

[e] Among those randomized to enhanced Internet plus telephone counseling (n=33 among TQD 0 days; n=77 among TQD 1-14 days, n=38 among TQD 15-28 days, and n=47 among TQD 29+ days).

[f] Participants were able to refuse answering a question or respond “I don’t know”. Sample sizes are follows: 30-day ppa, 426; no. quit attempts, 425.

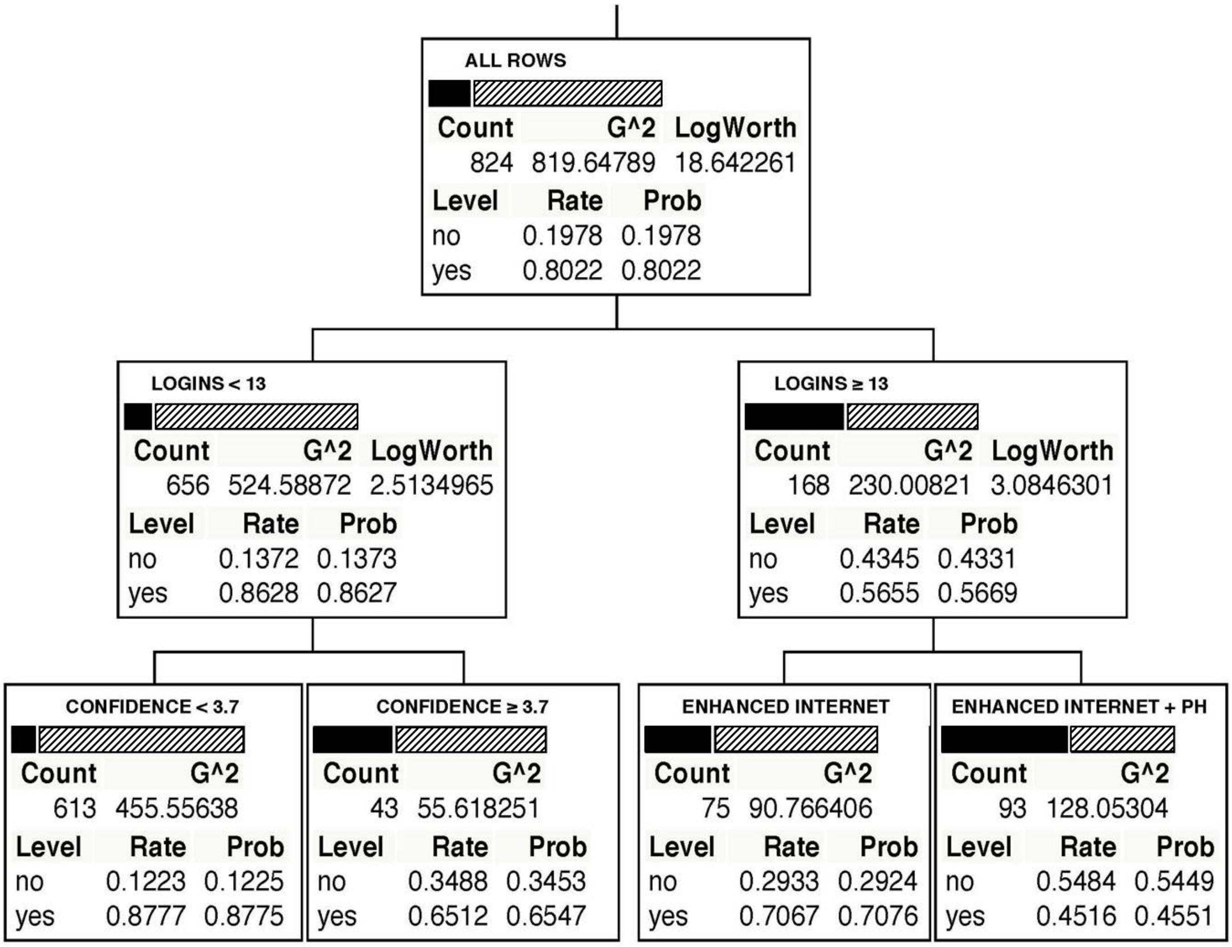
Study Question 4

The final logistic regression model among prospective quit date setters did not include desire to quit and confidence in quitting measures as both were unrelated to 30-day ppa. For 30-day ppa, the interaction between number of quit dates set and logins was significant (parameter estimate=–0.003, standard error=0.001, *P*=.005). Among those with high levels of website utilization (n=214; median split), 41.8% (33/79) of those who set one quit date were abstinent compared to 25.9% (35/135) of those who changed their quit date one or more times. Among those with high logins who set only one quit date and who were abstinent, the majority (60.6%, 20/33) opted to quit within 2 weeks of website registration.

CART Analysis

The CART model for 30-day ppa produced a tree with splits at three nodes (Figure 2), none of which were variables associated with quit date setting or timing. The first node, representing the total sample (n=824; 1 case missing outcome data), shows the overall proportion quit (19.8%, 163/824; Level=no) compared to the proportion smoking (80.2%, 661/824; Level=yes). The first split partitioned the total sample by logins (<13 logins, 79.6% of sample; ≥13 logins, 20.4% of sample). Among those who logged in <13 times, 13.7% (90/656) were abstinent, and for individuals who logged in ≥13 times, 43.5% (73/168) were abstinent. The second split occurred among individuals who logged in <13 times and was based on the Confidence Inventory scale score. Among those with a score <3.7, 12.2% (75/613) were abstinent compared to 34.9% (15/43) among those with a score ≥3.7. The third split occurred for those who logged in ≥13 times, where the sample was divided by treatment. Among enhanced Internet participants, 29.3% (22/75) were abstinent compared to 54.8% (51/93) of enhanced Internet plus telephone counseling participants. The k-fold cross-validation results showed good agreement (similar R squared values) between the folded and overall samples.

Figure 2. CART model predicting 30-day point prevalence abstinence at 3 months (n=824). Bars correspond to smoking status: solid=abstinent; diagonal lines=not abstinent. Count=total number of participants in subset; Level=smoking abstinence status (no/yes); Rate=relative proportion of the count in each abstinence status group; Logins=frequency of website logins during first 3 months of study; enhanced Internet and enhanced Internet + ph (telephone counseling)=treatment arms; Confidence=Confidence Inventory score.



Discussion

Principal Findings

In these secondary analyses of The iQUITT Study, we explored the use of an online interactive tool to set a quit date and its relationship to smoking outcomes. The majority of study participants who used the website set a quit date using the Quit Date Wizard: most set a prospective quit date, but some used it to document a quit date that had already passed. Only 9.7% (62/638) of those who used the website did not use this tool.

Our a priori hypotheses were only partially supported. We did not find evidence that individuals whose first TQD was set within 2 weeks of registration differed on baseline desire to quit or motivation to quit as hypothesized, but we did find that those who set a quit date within 2 weeks of registration had higher levels of baseline self-efficacy (Confidence Inventory score) and education compared to smokers who set later quit dates. We also found that participants whose TQD occurred within the first 2 weeks of website registration exhibited higher rates of website utilization than those with later quit dates. We did not find any differences in smoking outcomes based on latency to TQD. There was an interaction between website utilization and number of TQDs set on quit rates. At low levels of website utilization, there was no difference in abstinence rates based on number of quit dates set, but at high levels of website utilization, those who set only one quit date had significantly higher quit rates than those who changed their quit date.

Overall, the CART analysis was consistent with these findings. Latency to TQD did not predict abstinence, but website utilization (logins) and baseline self-efficacy did along with treatment group. Login frequency initially split the sample, and among individuals who logged into the website more frequently, the addition of telephone counseling appeared to increase abstinence relative to enhanced Internet alone. Self-efficacy appeared to be a key variable among those with lower levels of website utilization. Among this group, higher self-efficacy scores were associated with higher quit rates. This finding is consistent with a wealth of research demonstrating the importance of self-efficacy on smoking outcomes [52-55]. The importance of logins is consistent with other Web-based trials that have reported that website utilization is an important predictor of abstinence [31,56,57]. It should also be noted that none of the metrics of motivation to quit emerged in the CART, suggesting that website utilization was a stronger predictor of abstinence than motivation to quit.

In terms of the practical relevance of these results, both traditional and exploratory analyses both point to self-efficacy and website utilization as critical components of abstinence. Findings related to Study Questions 3 and 4 suggest that individuals who set a quit date early in the course of Web-based cessation treatment are more likely to be confident about their ability to achieve cessation and that setting a TQD early on and maintaining high levels of website utilization may incur an advantage for cessation. Taken together, these results suggest that Internet cessation programs should emphasize the importance of feeling efficacious about handling specific smoking situations and engaging with treatment at the highest level possible while potentially placing less emphasis on an absolute time frame for setting a TQD (ie, within 2 weeks versus 2-4 weeks). These results are consistent with a growing body of literature demonstrating the critical importance of engagement and adherence with regard to the effectiveness of Web-based health behavior change interventions [58-66].

Strengths and Limitations

These findings should be considered in the context of several related strengths and limitations. First, the CART analysis is a novel contribution to the literature concerning predictors of smoking abstinence. It is a powerful exploratory technique that offers an unbiased assessment of a large set of predictors and requires little input from the analyst. However, inferences based upon these analyses should be tested and replicated under controlled conditions. Second, participants were not required to set a quit date and could use the website as they desired, resulting in relatively naturalistic observations of the use of a quit date tool. Future research should examine how the use of this tool corresponds to self-reported quit attempts using other assessment methods. Third, we are unclear what to make of retrospective TQDs since current smoking status was confirmed during the baseline telephone survey. We speculate that participants may have used the Quit Date Wizard to document their most recent quit attempt or perhaps entered an erroneous date. Relative to registration, 55% of retrospective dates occurred within the week prior to study randomization, which suggests that many smokers search for cessation assistance in the early days following a quit attempt when they have returned to smoking. Qualitative methods or formal usability testing may shed light on this finding. Fourth, while the use of responder-only analyses is less conservative than intention-to-treat analyses, we feel this approach was appropriate for these exploratory analyses since imputation of missing data using an intent-to-treat (missing=smoking) approach might have obscured results. Fifth, it was not feasible to biochemically verify self-reported abstinence outcomes since this was a national sample recruited entirely via the Internet. Self-reported abstinence is a commonly accepted outcome metric in Web-based cessation trials [67-71] where misreporting of abstinence is expected to be minimal [72]. Last, we cannot rule out the possibility that low levels of website utilization were a consequence (and not cause) of relapse [73]. Studies that establish a chronological sequence of patterns of treatment utilization and relapse are needed [74].

Conclusions

In the context of a pragmatic randomized trial of Internet and telephone treatment for cessation, the timing of a TQD was not a significant predictor of cessation outcomes. Self-efficacy and an apparent commitment to an initial TQD were the components most highly related to abstinence but only via interactions with website utilization. Increasing treatment engagement has been noted as an important area for future research in Web-based cessation studies [57,75].

Acknowledgments

Primary funding for this work was from the National Cancer Institute at the National Institutes of Health (R01 CA104836). The funding agency had no involvement in the conduct of the study or preparation of this manuscript. There are no contractual constraints on publishing imposed by the funder. We thank Dr Lorien Abroms for her comments on an earlier version of this manuscript.

Conflicts of Interest

All authors are employees of Legacy, a nonprofit public health foundation that runs the BecomeAnEX website, an online tobacco cessation intervention.

Multimedia Appendix 1

Baseline characteristics of individuals who used the Quit Date Wizard (QD) (n=576) and those who did not (n=62).

[PDF File (Adobe PDF File), 8KB - jmir_v16i2e55_app1.pdf]

References

1. Abrams DB, Niaura R, Brown RA, Emmons KM, Goldstein MG, Monti PM. The tobacco dependence treatment handbook: A guide to best practices. New York, NY: The Guilford Press; 2003.
2. Fiore M, Jaén C, Baker T. Treating tobacco use and dependence: 2008 update. Clinical Practice Guideline. Rockville, MD: US Department of Health and Human Services, Public Health Service; May 2008.
3. McEwen AHP, Hajek P, McRobbie H, West RH. Manual of Smoking Cessation: A Guide for Counsellors And Practitioners. Oxford, UK: Blackwell Publishers; 2006.
4. Perkins KA, Conklin CA, Levine MD. Cognitive-Behavioral Therapy for Smoking Cessation: A Practical Guidebook to the Most Effective Treatments. New York, NY: Routledge; 2007.
5. Hallaq JH. The pledge as an instrument of behavioral change. J Soc Psychol 1976 Feb;98(First Half):147-148. [doi: 10.1080/00224545.1976.9923382] [Medline: 1271769]
6. Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement: A meta-analysis of effects and processes. Adv Exp Soc Psychol 2006;38:69-119. [doi: 10.1016/S0065-2601(06)38002-1]
7. Gollwitzer PM, Schaal B. Metacognition in action: The importance of implementation intentions. Pers Soc Psychol Rev 1998;2(2):124-136. [doi: 10.1207/s15327957pspr0202_5] [Medline: 15647140]
8. Shiffman S, Ferguson SG. Nicotine patch therapy prior to quitting smoking: A meta-analysis. Addiction 2008 Apr;103(4):557-563. [doi: 10.1111/j.1360-0443.2008.02138.x] [Medline: 18339101]
9. Balmford J, Borland R, Burney S. The influence of having a quit date on prediction of smoking cessation outcome. Health Educ Res 2010 Aug;25(4):698-706 [FREE Full text] [doi: 10.1093/her/cyq013] [Medline: 20194359]
10. Segan CJ, Borland R, Greenwood KM. Do transtheoretical model measures predict the transition from preparation to action in smoking cessation? Psychology & Health 2002 Jan;17(4):417-435. [doi: 10.1080/0887044022000004911]
11. de Vries H, Eggers SM, Bolman C. The role of action planning and plan enactment for smoking cessation. BMC Public Health 2013;13:393 [FREE Full text] [doi: 10.1186/1471-2458-13-393] [Medline: 23622256]
12. Etter JF. Comparing abrupt and gradual smoking cessation: A randomized trial. Drug Alcohol Depend 2011 Nov 1;118(2-3):360-365. [doi: 10.1016/j.drugalcdep.2011.04.016] [Medline: 21571448]
13. Ferguson SG, Shiffman S, Gitchell JG, Sembower MA, West R. Unplanned quit attempts--results from a U.S. sample of smokers and ex-smokers. Nicotine Tob Res 2009 Jul;11(7):827-832. [doi: 10.1093/ntr/ntp072] [Medline: 19509277]
14. Larabie LC. To what extent do smokers plan quit attempts? Tob Control 2005 Dec;14(6):425-428 [FREE Full text] [doi: 10.1136/tc.2005.013615] [Medline: 16319368]
15. Sendzik T, McDonald PW, Brown KS, Hammond D, Ferrence R. Planned quit attempts among Ontario smokers: impact on abstinence. Addiction 2011 Nov;106(11):2005-2013. [doi: 10.1111/j.1360-0443.2011.03498.x] [Medline: 21592249]
16. West R, Sohal T. "Catastrophic" pathways to smoking cessation: findings from national survey. BMJ 2006 Feb 25;332(7539):458-460 [FREE Full text] [doi: 10.1136/bmj.38723.573866.AE] [Medline: 16443610]
17. Murray RL, Lewis SA, Coleman T, Britton J, McNeill A. Unplanned attempts to quit smoking: missed opportunities for health promotion? Addiction 2009 Nov;104(11):1901-1909. [doi: 10.1111/j.1360-0443.2009.02647.x] [Medline: 19681806]
18. Hughes JR, Callas PW. Is delaying a quit attempt associated with less success? Nicotine Tob Res 2011 Dec;13(12):1228-1232 [FREE Full text] [doi: 10.1093/ntr/ntr207] [Medline: 21908460]
19. Hughes JR, Russ C, Messig MA. Association of deferring a quit attempt with smoking cessation success: a secondary analysis. J Subst Abuse Treat 2014 Feb;46(2):264-267. [doi: 10.1016/j.jsat.2013.08.015] [Medline: 24074849]
20. American Cancer Society. How to quit URL: http://www.cancer.org/healthy/stayawayfromtobacco/guidetoquittingsmoking/guide-to-quitting-smoking-how-to-quit [accessed 2013-06-16] [WebCite Cache ID 6HQWmRP5s]
21. American Legacy Foundation. Set your quit date URL: http://www.becomeanex.org/set-your-quit-date.php [accessed 2013-06-16] [WebCite Cache ID 6HQWxmHma]
22. Smokefree.gov. Quit guide: Preparing to quit URL: http://www.smokefree.gov/qg-preparing-set.aspx [WebCite Cache ID 6HQX6fwW0]
23. Altria Client Services. Get Ready URL: http://www.quitassist.com/en/cms/five-keys-for-quitting/Get-Ready/default.aspx [accessed 2013-06-16] [WebCite Cache ID 6HQXH04Th]
24. 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]
25. Fox S. Pew Research Center. 2005. Health information online URL: http://www.pewinternet.org/~/media//Files/Reports/2005/PIP_Healthtopics_May05.pdf.pdf [WebCite Cache ID 6IztI1kpM]
26. Fox S. 2006. Online health search 2006 URL: http://www.pewinternet.org/~/media//Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf [accessed 2013-08-19] [WebCite Cache ID 6Izt6K56K]

27. Rennard S, Hughes J, Cinciripini PM, Kralikova E, Raupach T, Arteaga C, Flexible Quit Date Study Group. A randomized placebo-controlled trial of varenicline for smoking cessation allowing flexible quit dates. Nicotine Tob Res 2012 Mar;14(3):343-350 [FREE Full text] [doi: 10.1093/ntr/ntr220] [Medline: 22080588]
28. Borrelli B, Papandonatos G, Spring B, Hitsman B, Niaura R. Experimenter-defined quit dates for smoking cessation: adherence improves outcomes for women but not for men. Addiction 2004 Mar;99(3):378-385. [doi: 10.1111/j.1360-0443.2004.00648.x] [Medline: 14982551]
29. Borland R, Partos TR, Cummings KM. Systematic biases in cross-sectional community studies may underestimate the effectiveness of stop-smoking medications. Nicotine Tob Res 2012 Dec;14(12):1483-1487 [FREE Full text] [doi: 10.1093/ntr/nts002] [Medline: 22318689]
30. 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]
31. 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]
32. 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 [FREE Full text] [doi: 10.1001/archinternmed.2010.451] [Medline: 21220660]
33. Breiman L, Freideman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Belmont, CA: Wadsworth International Group; 1984.
34. Kitsantas P, Moore TW, Sly DF. Using classification trees to profile adolescent smoking behaviors. Addict Behav 2007 Jan;32(1):9-23. [doi: 10.1016/j.addbeh.2006.03.014] [Medline: 16677775]
35. Zabor EC, Li Y, Thornton LM, Shuman MR, Bulik CM, Lichtenstein P, et al. Initial reactions to tobacco use and risk of future regular use. Nicotine Tob Res 2013 Feb;15(2):509-517 [FREE Full text] [doi: 10.1093/ntr/nts180] [Medline: 22949572]
36. 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 [FREE Full text] [Medline: 17491170]
37. Quitnet.com. Quitnet homepage URL: http://www.quitnet.com/qnhomepage.aspx [accessed 2013-11-04] [WebCite Cache ID 6KsxaJn3E]
38. 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 [FREE Full text] [doi: 10.1080/14622200500055319] [Medline: 16036277]
39. Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. Am J Public Health 2010 Jul;100(7):1282-1289 [FREE Full text] [doi: 10.2105/AJPH.2009.165449] [Medline: 20466971]
40. Zhu SH, Stretch V, Balabanis M, Rosbrook B, Sadler G, Pierce JP. Telephone counseling for smoking cessation: effects of single-session and multiple-session interventions. J Consult Clin Psychol 1996 Feb;64(1):202-211. [Medline: 8907100]
41. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
42. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [Medline: 1932883]
43. 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]
44. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
45. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 1994;10(2):77-84. [Medline: 8037935]
46. Borrelli B, Mermelstein R. The role of weight concern and self-efficacy in smoking cessation and weight gain among smokers in a clinic-based cessation program. Addict Behav 1998;23(5):609-622. [Medline: 9768298]
47. Cohen S, Doyle WJ, Skoner DP, Rabin BS, Gwaltney JM. Social ties and susceptibility to the common cold. JAMA 1997 Jun 25;277(24):1940-1944. [Medline: 9200634]
48. Graham AL, Papandonatos GD, Bock BC, Cobb NK, Baskin-Sommers A, Niaura R, et al. Internet- vs. telephone-administered questionnaires in a randomized trial of smoking cessation. Nicotine Tob Res 2006 Dec;8 Suppl 1:S49-S57 [FREE Full text] [Medline: 17491171]
49. Cohen S, Lichtenstein E. Partner behaviors that support quitting smoking. J Consult Clin Psychol 1990 Jun;58(3):304-309. [Medline: 2365893]
50. Brown RL, Leonard T, Saunders LA, Papasouliotis O. A two-item conjoint screen for alcohol and other drug problems. J Am Board Fam Pract 2001;14(2):95-106 [FREE Full text] [Medline: 11314930]
51. Gaudard M, Ramsey P, Stephens M. Interactive data mining and design of experiments: The JMP partition and custom design platforms. Brookline, New Hampshire: North Haven Group; 2006 Mar. URL: http://www.jmp.com/software/whitepapers/pdfs/372455_interactive_datamining.pdf [accessed 2014-02-11] [WebCite Cache ID 6NJ1kn3mT]

52. O'Hea EL, Boudreaux ED, Jeffries SK, Carmack Taylor CL, Scarinci IC, Brantley PJ. Stage of change movement across three health behaviors: the role of self-efficacy. Am J Health Promot 2004;19(2):94-102. [Medline: 15559709]
53. Prochaska JO, Crimi P, Lapsanski D, Martel L, Reid P. Self-change processes, self-efficacy and self-concept in relapse and maintenance of cessation of smoking. Psychol Rep 1982 Dec;51(3 Pt 1):983-990. [Medline: 7163454]
54. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]
55. Bandura A. Self-efficacy: The exercise of control. New York: W.H. Freeman; 1997.
56. Saul JE, Schillo BA, Evered S, Luxenberg MG, Kavanaugh A, Cobb N, et al. Impact of a statewide Internet-based tobacco cessation intervention. J Med Internet Res 2007;9(3):e28 [FREE Full text] [doi: 10.2196/jmir.9.4.e28] [Medline: 17942390]
57. Richardson A, Graham AL, Cobb N, Xiao H, Mushro A, Abrams D, et al. Engagement promotes abstinence in a Web-based cessation intervention: cohort study. J Med Internet Res 2013;15(1):e14 [FREE Full text] [doi: 10.2196/jmir.2277] [Medline: 23353649]
58. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
59. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
60. Wangberg SC, Bergmo TS, Johnsen JA. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
61. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. Strategies to facilitate exposure to Internet-delivered health behavior change interventions aimed at adolescents or young adults: a systematic review. Health Educ Behav 2011 Feb;38(1):49-62. [doi: 10.1177/1090198110372878] [Medline: 21189422]
62. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to Web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
63. Hutton HE, Wilson LM, Apelberg BJ, Tang EA, Odelola O, Bass EB, et al. A systematic review of randomized controlled trials: Web-based interventions for smoking cessation among adolescents, college students, and adults. Nicotine Tob Res 2011 Apr;13(4):227-238. [doi: 10.1093/ntr/ntq252] [Medline: 21350042]
64. 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]
65. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and Internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2 Suppl):TC9-T20. [doi: 10.1097/NCN.0b013e3182155274] [Medline: 21372645]
66. 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]
67. 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 [FREE Full text] [doi: 10.2196/jmir.1605] [Medline: 22169631]
68. Leykin Y, Aguilera A, Torres LD, Pérez-Stable EJ, Muñoz RF. Interpreting the outcomes of automated Internet-based randomized trials: example of an international smoking cessation study. J Med Internet Res 2012;14(1):e5 [FREE Full text] [doi: 10.2196/jmir.1829] [Medline: 22314016]
69. 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 [FREE Full text] [doi: 10.1016/j.amepre.2007.12.024] [Medline: 18407003]
70. Rabius V, Pike KJ, Wiatrek D, McAlister AL. Comparing Internet assistance for smoking cessation: 13-month follow-up of a six-arm randomized controlled trial. J Med Internet Res 2008;10(5):e45 [FREE Full text] [doi: 10.2196/jmir.1008] [Medline: 19033146]
71. 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]
72. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002 May;4(2):149-159. [doi: 10.1080/14622200210123581] [Medline: 12028847]
73. Raupach T, Brown J, Herbec A, Brose L, West R. A systematic review of studies assessing the association between adherence to smoking cessation medication and treatment success. Addiction 2014 Jan;109(1):35-43. [doi: 10.1111/add.12319] [Medline: 23919621]

74. Borland R, Balmford J, Swift E. Effects of timing of initiation and planning on smoking cessation outcomes: study protocol for a randomised controlled trial. BMC Public Health 2013;13:235 [FREE Full text] [doi: 10.1186/1471-2458-13-235] [Medline: 23496992]
75. Graham AL, Cha S, Papandonatos GD, Cobb NK, Mushro A, Fang Y, et al. Improving adherence to Web-based cessation programs: a randomized controlled trial study protocol. Trials 2013;14:48 [FREE Full text] [doi: 10.1186/1745-6215-14-48] [Medline: 23414086]

Abbreviations

ANOVA: analysis of variance
CART: classification and regression tree
CES-D: Center for Epidemiologic Studies-Depression Scale
ph: telephone counseling
ppa: point prevalence abstinence
QD: quit date
TQD: target quit date

Edited by G Eysenbach; submitted 06.11.13; peer-reviewed by J McClure, M Brunette; comments to author 28.11.13; revised version received 09.12.13; accepted 28.01.14; published 17.02.14.

Please cite as:
Cobb CO, Niaura RS, Donaldson EA, Graham AL
Quit Now? Quit Soon? Quit When You're Ready? Insights About Target Quit Dates for Smoking Cessation From an Online Quit Date Tool
J Med Internet Res 2014;16(2):e55
URL: http://www.jmir.org/2014/2/e55/
doi:10.2196/jmir.3086
PMID:24534139

Review

Digital Asthma Self-Management Interventions: A Systematic Review

Deborah Morrison[1], BSc, MB ChB; Sally Wyke[2], BSc, PhD; Karolina Agur[1], MD; Euan J Cameron[3], MB ChB, PhD; Robert I Docking[4], MB ChB; Alison M MacKenzie[4], MB ChB; Alex McConnachie[5], BSc, MSc, PhD; Vandana Raghuvir[6]; Neil C Thomson[3], MB ChB, MD; Frances S Mair[1], MB ChB, MD

[1]General Practice & Primary Care, Institute of Health & Wellbeing, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, United Kingdom

[2]Institute of Health & Wellbeing, College of Social Sciences, University of Glasgow, Glasgow, United Kingdom

[3]Respiratory Medicine, Institute of Infection, Immunity & Inflammation, University of Glasgow, Glasgow, United Kingdom

[4]Forth Valley Royal Hospital, NHS Forth Valley, United Kingdom

[5]Robertson Centre for Biostatistics, Institute of Health & Wellbeing, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, United Kingdom

[6]School of Medicine, University of Glasgow, Glasgow, United Kingdom

Corresponding Author:
Frances S Mair, MB ChB, MD
General Practice & Primary Care
Institute of Health & Wellbeing, College of Medical, Veterinary & Life Sciences
University of Glasgow
1 Horselethill Road
Glasgow, G12 9LX
United Kingdom
Phone: 44 (0)141 330 8312
Fax: 44 (0)141 330 8332
Email: frances.mair@glasgow.ac.uk

Abstract

Background: Many people with asthma tolerate symptoms and lifestyle limitations unnecessarily by not utilizing proven therapies. Better support for self-management is known to improve asthma control, and increasingly the Internet and other digital media are being used to deliver that support.

Objective: Our goal was to summarize current knowledge, evidenced through existing systematic reviews, of the effectiveness and implementation of digital self-management support for adults and children with asthma and to examine what features help or hinder the use of these programs.

Methods: A comprehensive search strategy combined 3 facets of search terms: (1) online technology, (2) asthma, and (3) self-management/behavior change/patient experience. We undertook searches of 14 databases, and reference and citation searching. We included qualitative and quantitative systematic reviews about online or computerized interventions facilitating self-management. Title, abstract, full paper screening, and quality appraisal were performed by two researchers independently. Data extraction was undertaken using standardized forms.

Results: A total of 3810 unique papers were identified. Twenty-nine systematic reviews met inclusion criteria: the majority were from the United States (n=12), the rest from United Kingdom (n=6), Canada (n=3), Portugal (n=2), and Australia, France, Spain, Norway, Taiwan, and Greece (1 each). Only 10 systematic reviews fulfilled pre-determined quality standards, describing 19 clinical trials. Interventions were heterogeneous: duration of interventions ranging from single use, to 24-hour access for 12 months, and incorporating varying degrees of health professional involvement. Dropout rates ranged from 5-23%. Four RCTs were aimed at adults (overall range 3-65 years). Participants were inadequately described: socioeconomic status 0/19, ethnicity 6/19, and gender 15/19. No qualitative systematic reviews were included. Meta-analysis was not attempted due to heterogeneity and inadequate information provision within reviews. There was no evidence of harm from digital interventions. All RCTs that examined knowledge (n=2) and activity limitation (n=2) showed improvement in the intervention group. Digital interventions improved markers of self care (5/6), quality of life (4/7), and medication use (2/3). Effects on symptoms (6/12) and school absences

(2/4) were equivocal, with no evidence of overall benefits on lung function (2/6), or health service use (2/15). No specific data on economic analyses were provided. Intervention descriptions were generally brief making it impossible to identify which specific "ingredients" of interventions contribute most to improving outcomes.

Conclusions: Digital self-management interventions show promise, with evidence of beneficial effects on some outcomes. There is no evidence about utility in those over 65 years and no information about socioeconomic status of participants, making understanding the "reach" of such interventions difficult. Digital interventions are poorly described within reviews, with insufficient information about barriers and facilitators to their uptake and utilization. To address these gaps, a detailed quantitative systematic review of digital asthma interventions and an examination of the primary qualitative literature are warranted, as well as greater emphasis on economic analysis within trials.

(J Med Internet Res 2014;16(2):e51) doi:10.2196/jmir.2814

KEYWORDS

asthma; self-management; Internet; eHealth; systematic review; patient education

Introduction

Asthma is common, affecting an estimated 300 million people worldwide. The number of disability-adjusted life years lost is estimated at 15 million per year, similar to that for diabetes [1]. The main goals of treatment for asthma include achieving and maintaining control of symptoms, normal activity levels, minimal exacerbations, normal lung function, and preventing deaths from asthma [2]. However, these goals are not widely achieved; people with asthma often tolerate unnecessary symptoms, and management of the condition can often be suboptimal [1,3,4]. Guided self-management for asthma as part of systematic, planned care can lead to improvements in patient outcomes such as increases in knowledge, confidence to manage asthma, and improved quality of life, as well as reductions in hospitalizations, emergency room visits, unscheduled visits to the doctor, and days off work or school [5-8].

Despite evidence of benefits, guided self-management, particularly through the use of asthma plans, remains underused [9-11]. While interventions can often be successful in trial settings, evidence of their implementation into every day practice is limited [9,12]. Therefore, there is growing interest in the potential of the Internet and other digital media as a medium to deliver more tailored, relevant self-management support, while maintaining cost-effectiveness, with greater scope for integration into the everyday lives of those with asthma.

While many reviews have been published in the field of self-management in asthma, there is a lack of clarity about the role of digital interventions and which specific components of interventions or "ingredients" contribute most to promoting effective self-management practices and translate into improvement in patient outcomes. There is increasing interest in standardizing methods of determining the "active ingredients" of self-management interventions, potentially making it easier to measure and reproduce those features found to be most effective in future interventions [13].

The aim of this paper is to summarize current knowledge, evidenced through existing systematic reviews, of the effectiveness and implementation of digital self-management support for adults and children with asthma and to examine what features help or hinder the use of these programs. We describe our metareview, which examines the effects, if any, of asthma digital self-management interventions on a range of measures of lung function, symptoms, quality of life, and health care utilization.

Methods

Overview

We conducted a systematic review of systematic reviews—an approach that has proven helpful in synthesizing a broad base of literature in order to identify research gaps and inform future intervention programs [14]. Our aims were to assess the evidence of effectiveness of asthma digital interventions for self-management as measured by an inclusive range of clinical and process outcomes.

We documented recruitment and retention rates, information about implementation processes, whether cost effectiveness was assessed, and whether theories of behavior change were used in intervention development to help us gain a better of understanding of features that helped or hindered the use of the programs. Our protocol is available in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

We included qualitative and quantitative reviews and used the PICOS (participants, interventions, comparison, outcomes, study design) framework [15] to define inclusion criteria (see Textbox 1 for details). Our exclusion criteria can be found in Textbox 2.

Textbox 1. Inclusion criteria.

- Participants: those with asthma of any age or their caregivers.
- Intervention: online or computerized interventions facilitating self-management through education and/or providing advice or other behavior change approach. We included only interventions that provided these features independent of any health professional input. Interventions delivered by computer, tablet, smartphone, or purpose-built electronic device were included.
- Comparison: usual care or other forms of self-management interventions, such as face-to-face education or written information.
- Primary outcomes:
  - activity limitation (eg, days off work/school/disturbed nights)
  - adverse events
  - barriers and facilitators to online asthma intervention use by patients and practitioners
  - biomarkers of airway inflammation (eg, exhaled nitric oxide)
  - health service utilization (including scheduled/unscheduled, and primary/secondary care)
  - lung function (eg, spirometry & reversibility, peak expiratory flow [PEF])
  - medication use (eg, relief inhaled β agonist use, compliance with medication)
  - quality of life
  - symptoms (measures of asthma control, eg, diary card scores, asthma control questionnaire, exacerbation rates)
- Secondary outcomes:
  - markers of self-management (eg, adherence to monitoring tools, use of action plans, self-efficacy)
  - patient knowledge
  - patient satisfaction
  - recruitment, retention rates
  - cost effectiveness
  - use of behavior change theory during intervention development and implementation processes
- Study design: Quantitative reviews describing randomized controlled trials (RCTs) and qualitative reviews seeking to understand the patients or providers' experience of using these asthma interventions and those that describe the theory behind the development of such interventions. The full definition of a systematic review used is found in Multimedia Appendix 1.

Textbox 2. Exclusion criteria.

- Intervention: interventions consisting only of telemonitoring or clinical decision support software for health professionals were excluded. Interventions that provided only a means of self-monitoring without direct feedback were excluded (eg, electronic diaries for recording peak flows or symptoms that did not provide automated feedback). The content of the intervention was required to be delivered at least in part by the digital medium itself. Devices that were simply digital modes of communicating between patients and health professionals were excluded.
- Outcomes: reviews that did not provide information specific to our outcomes of interest.
- Study design: conference proceedings and theses, and for quantitative reviews, non-RCTs were excluded.

Information Sources and Search Strategy

A professional systematic review company (York Health Economic Consortium) searched a wide range of databases covering health, mental health, education, social science (14 in total), with no start date before July 2011. The search strategy was devised using a combination of subject indexing terms (eg, MeSH [medical subject headings] in MEDLINE), and free-text search terms in the title and abstract. The search terms were identified through discussion between the research team, by scanning background literature, and by browsing a database's thesaurus. To ensure sensitivity, the search strategy did not include a methodological search filter to identify reviews. The searches were not limited by date range or language. Hand-searching of Patient Education and Counseling, and the Primary Care Respiratory Journal, and reference searching and citation searching of included studies was undertaken (DM). We contacted experts to establish if any reviews had been missed. The search of electronic databases was updated to October 3, 2013.

The search strategy covered 3 broad areas: (1) asthma and related terms, (2) online/computerized and related terms, and (3) self-care/self-management, patient experience, qualitative, and related terms. The full list of databases searched and an example of the full search strategy for MEDLINE are available in Multimedia Appendix 2.

Study Selection

Titles, abstracts, and full papers were screened by one researcher (DM) plus one other independent researcher (EC, SW, FM, NCT, KA, RD, AM, or VR). We independently undertook quality appraisal; disagreements were resolved by discussion with a third party if necessary. Only studies meeting predetermined quality criteria advanced to data extraction.

Data Collection

We used online data collection forms using Distiller SR software. For each included review, we collected (1) general information about the review (year, country of first author, language, number of studies, number of asthma studies, and number of digital asthma studies), (2) descriptions of each included RCT and intervention that fulfilled our criteria (inclusion criteria were applied to the systematic reviews initially and then subsequently to their featured RCTs, to determine those relevant to this review), and (3) results for each outcome of interest and the original article this result was derived from (including quotes from qualitative/narrative reviews).

Quality Appraisal

Quality appraisal was undertaken in two ways. First, at the full paper screening stage, papers were required to meet criteria laid out in our definition of a review (eg, evidence of a systematic search or criteria for selection of papers must be included; see Multimedia Appendix 1 for full definition). Then, the included papers underwent formal quality appraisal using A Measurement Tool to Assess Systematic Reviews (AMSTAR) [16-18]. This 11-point checklist covers the following domains: establishing the research question and inclusion criteria before the conduct of the review, data extraction by at least 2 independent data extractors, comprehensive literature review with searching of at least two databases, detailed list of included/excluded studies, quality assessment of included studies and consideration of quality assessments in analysis and conclusions, appropriate assessment of homogeneity, assessment of publication bias, and a statement of any conflict of interest. We made minor alterations to the wording in order to make the checklist applicable to non-quantitative reviews (available on request from the corresponding author). Papers needed to achieve at least 50% of quality indicators, plus a “yes” to question 7, which asks “Was the scientific quality of the included studies assessed and documented?” This was required to allow us to make a comment on the quality of the included data. Two researchers (DM + [KA, RD, AM, or VR]) scored each paper independently, with conflicts resolved by discussion with a third party if necessary (FM).

Any data provided describing risk of bias across reviews, and within individual RCTs, were extracted to inform the discussion.

Data Synthesis

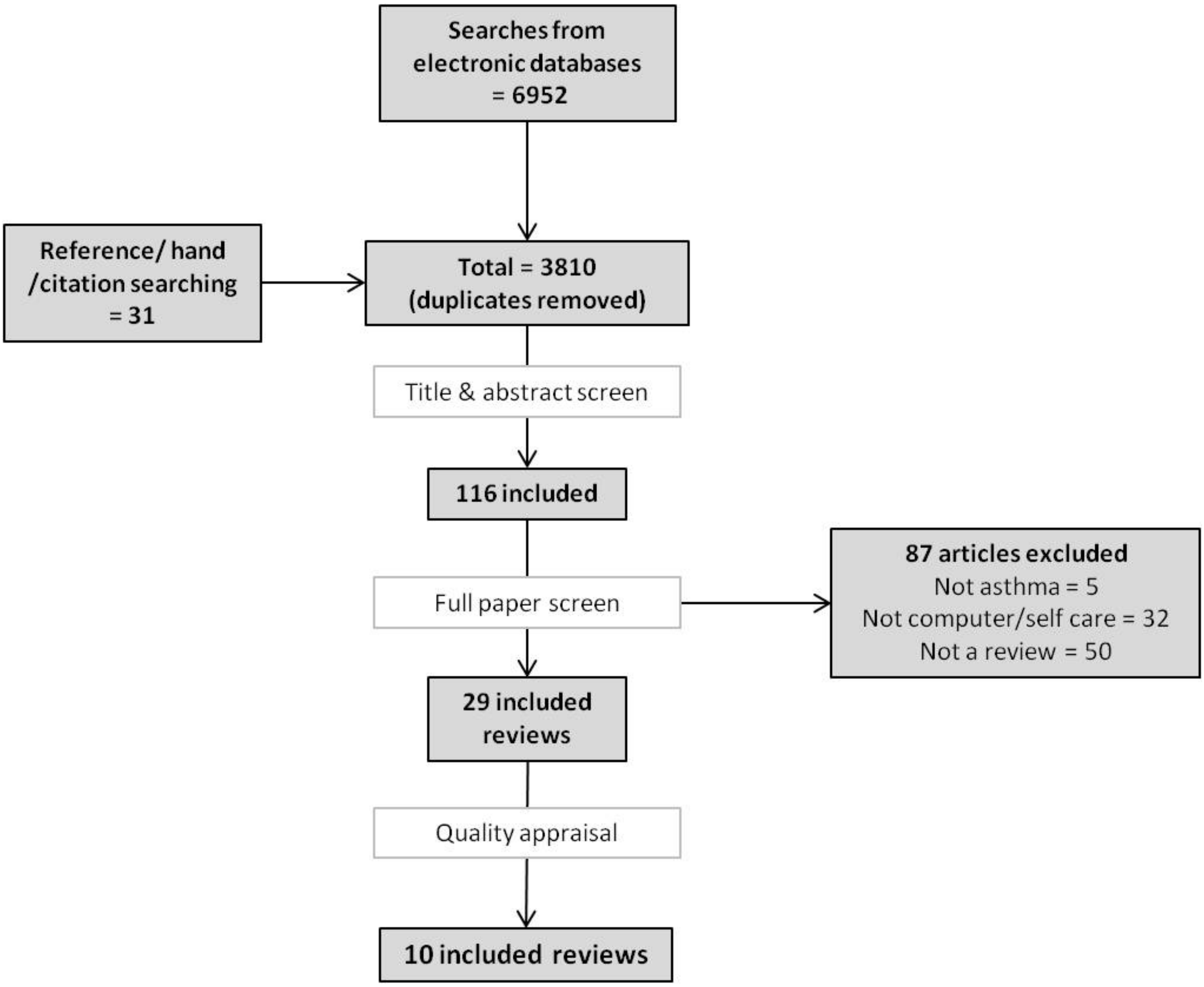
Meta-analysis was not possible. Quantitative results for individual outcomes results were described as favoring the intervention group, favoring the control group, or demonstrating no difference. All results were included in a narrative summary.

Results

Results of Article Screening and Selection

Our search identified 3810 unique citations, and title and abstract screening identified 116 full papers for review. Of these, 29 fulfilled our inclusion criteria (Figure 1).

Figure 1. Flow chart demonstrating the screening process of papers in the systematic review.



Description of Included Reviews

The 29 papers included systematic reviews from United States (12), United Kingdom (6), Canada (3), Portugal (2) Australia, France, Spain, Norway, Taiwan, and Greece (1 each). Only 9 reviews focused completely on asthma interventions (the others included self-management interventions for asthma alongside a range of other chronic conditions such as diabetes, heart failure, and hypertension), and 11 reviews looked exclusively at digital interventions (the rest examined digital interventions alongside other modes of self-management). These 29 reviews consisted of 4 Cochrane reviews [19-22] and 3 meta-analyses [23-25], with the remaining 22 being narrative or descriptive studies (see Table 1; [10,19-47]).

Table 1. Included reviews with quality appraisal results.

First author of systematic review	Year	RCTs, n	Asthma RCTs, n	Digital asthma RCTs, n	Country, language	AMSTAR, %	Question 7=Yes[a]
Krishna et al [26]	1997	22	2	2	United States, English	10	N
Lewis [27]	2003	32	7	4	United States, English	40	N
Wantland et al [23]	2004	20	2	2	United States, English	36	N
Murray et al [19][b]	2005	24	6	5	United Kingdom, English	82	Y
Almeida et al [28]	2006	13	13	12	Portugal, English	11	N
Bussey-Smith & Rossen [29]	2007	9	9	9	United States, English	30	N
Garcia-Lizana & Sarria-Santamera [30]	2007	24	5	5	Spain, English	40	N
Ring et al [10][b]	2007	14	14	1	United Kingdom, English	70	Y
Coffman [25]	2008	37	37	6	United States, English	20	N
Fox [31]	2009	25	4	4	United States, English	10	N
Moeinedin et al [32]	2009	27	8	7	Canada, English	40	Y
Papastergiou [33]	2009	34	3	3	Greek, English	20	N
Stinson et al [34][b]	2009	9	5	3	Canada, English	70	Y
Boyd et al [20][b]	2009	30	30	2	Australia, English	90	Y
Coscrato et al [35]	2010	16	3	3	Portugal, Portuguese	30	Y
Cushing & Steele [24]	2010	33	9	9	United States, English	45	N
Gremeaux et al [36]	2010	39	4	4	France, English	11	N
McDermott & While [37][b]	2010	15	2	2	United Kingdom, English	50	Y
McLean et al [22][b]	2010	21	21	5	United Kingdom, English	100	Y
Pare et al [38]	2010	62	8	1	Canada, English	40	Y
Klasnja et al [39]	2011	n/a[c]	n/a	n/a	United States, English	10	N
Welsh et al [21][b]	2011	12	12	1	United Kingdom, English	90	Y
Chia-Chi Kuo & Hsiu-Hung Wang [40]	2012	12	4	1	Taiwan, Chinese	30	Y
Chrisler [41]	2012	18	18	6	United States, English	20	N
Johansen et al, (2 part review) [42,43][b]	2012	29	7	5	Norway, English	60	Y
Kirk et al [44][b]	2012	13	10	2	United Kingdom, English	50	Y
Mosnaim et al [45]	2012	17	17	10	United States, English	20	N
Nickels & Dimov [46]	2012	3	2	2	United States, English	10	N
Hieftje et al [47]	2013	19	6	6	United States, English	50	Y

[a] AMSTAR Q7 relates to whether the review assesses and documents any quality appraisal of their included studies.

[b] These studies meet appraisal criteria.

[c] Klasnja et al was a qualitative paper and did not specifically provide details of included papers.

Quality of the Included Reviews

Within the 29 included reviews, descriptions of the included RCTs and particularly the interventions themselves were generally suboptimal. AMSTAR scores ranged from 10% to 100% with a median of 40% (interquartile range 20–50), with 10 reviews scoring 50% or greater. All reviews with 50% or over also provided an individual assessment of quality, which was an essential criterion to progress to data extraction. Areas where reviews performed particularly poorly include providing information about conflict of interests of included RCTs (1/29), providing a priori design (6/29), independent screening and data extraction (6/29 clearly demonstrated this; a further 15/29 may have, but insufficient information was provided to be able to confirm this), and providing a list of excluded studies (7/29).

Of the 10 reviews meeting the predetermined AMSTAR requirements (Table 1), these included the 4 Cochrane reviews and 6 narrative/descriptive reviews. Four reviews focused only on asthma, with the remaining 6 featuring a range of conditions. Four of the reviews included nondigital/electronic modes of delivery of interventions. There was no one review that looked specifically at digital interactive interventions aimed only at those with asthma. These 10 systematic reviews presented results from a total of 19 RCTs meeting our inclusion criteria (see Table 2 for descriptions [10,19-22,34,37,42-44,47-68] and Table 3 for interventions [48-66]).

All 10 systematic reviews provided sample sizes for their included RCTs, but only three provided numbers allocated to intervention group versus control [19,20,37]. All reviews presented some information about age, 8 out of 10 presented information about gender, and only 4 of the 10 reviews about the ethnicity of participants [19-21,44]. No review provided information about socioeconomic status or levels of educational attainment.

Table 2. Description of included RCTs within included systematic reviews.

Systematic review	Author (year of RCT)	Sample size, n	Age, years	% male[a]	Ethnicity[b]	Nos. at follow-up	Drop-out rate, %	Duration	Country
McLean 2010 [22]	Cruz-Correia 2007 [48]	21	16-65	n/a	n/a	n/a	n/a	4 wks intervention, 4 wks control	Portugal
	Guendelman 2002 [49]	134	8-16	n/a	n/a	122/134	9	12 wks	United States
	Jan 2007 [50]	196	6-12	n/a	n/a	153/196	22	12 wks	Taiwan
	Rasmussen 2005 [51]	300	18-45	n/a	n/a	253/300	16	52 wks	Denmark
	Van der Meer 2009 [52]	200	18-50	n/a	n/a	183/200	8	52 wks	Netherlands
Stinson 2009 [34]	Jan 2007 [50]	n/a[c]	6-12	38	n/a	n/a	17	12 wks	Taiwan
	Joseph 2007 [53]	n/a[c]	5-19	37	n/a	n/a	17	26 wks	United States
	Krishna 2003 [54]	n/a[c]	7-17	65	n/a	n/a	10	52 wks	United States
Ring 2007 [10]	Rasmussen 2005 [51]	300	18-45	32 34[d]	n/a	253/300	16	n/a	Denmark
Murray 2005[e][19]	Bartholomew 2000 [55]	171	7-17	65	Hispanic 42%, AA 53%	133/171	22	4 to 15.6 months, mean 7.6 months	United States
	Guendelman 2002 [49]	134	8-16	57	AA 76%	120/134	10	12 wks	United States
	Homer 2000 [56]	137	3-12	69	AA 60.5%, Hispanic 5.3%	106/137	23	Used over 3 sessions at community clinic	United States
	Krishna 2003 [54]	246	7-17	65	White 86%	228/246	7	Approx. 80 minutes to complete. Used during 3 routine clinic visits.	United States
	Shegog 2001 [57]	76	10.7 (mean)	61	White 48%, AA 41%, Hispanic 7%	71/76	7	Game played during session a medical centre	United States
Hieftje 2013 [47]	Bartholomew 2000 [55]	133	7-17	65	n/a	n/a	n/a	15.6 months	n/a
	HussB 2003[f][58]	148	7-12	44	n/a	n/a	n/a	12 wks	n/a
	McPherson 2006 [59]	101	7-14	53	n/a	n/a	n/a	6 months	n/a
	Rubin 1986 [60]	65	7-12	n/a	n/a	n/a	n/a	10 months	n/a
	Shames 2004 [61]	119	5-12	60	n/a	n/a	n/a	12 months	n/a
	Vilozni 2001 [62]	112	3-6	44	n/a	n/a	n/a	One-off session	n/a
Boyd 2010 [20]	Homer 2000 [56]	137	3-12	69	AA 61%	106/137	23	12 wks (used game 3 times)	United States
	Shames 2004 [61]	119	5-12	58	Hispanic 57%, AA 21%	97/119	18	32 wks	United States

Systematic review	Author (year of RCT)	Sample size, n	Age, years	% male[a]	Ethnicity[b]	Nos. at follow-up	Drop-out rate, %	Duration	Country
Johansen 2012 [42,43]	Chan 2007 [63]	120	6-17	63	n/a	n/a	n/a	6 wks (follow-up 12 months)	United States
	Guendelman 2002 [49]	134	8-16	40 (i) 37 (c)	n/a	n/a	n/a	Follow-up 3 months	United States
	Jan 2007 [50]	164	6-12	40 (i) 37 (c)	n/a	n/a	n/a	Follow-up 3 months	Taiwan
	Rasmussen 2005 [51]	300	18-45	31	n/a	n/a	n/a	Follow up 6 months	Denmark
	Van der Meer 2009 [52]	200	18-50	31	n/a	n/a	n/a	Follow-up 12 months	Netherlands
Welsh 2011 [21]	Kamps 2004[g][64]	20	7-12	57 (i) 75 (c)	AA 20%, European American 53%, Hispanic American 27%	15	25	6 wks	United States
McDermott 2013 [37]	Sundberg 2005 [65]	97	18-25	n/a	n/a	n/a	n/a	One 1 hour intervention, follow-up at 1 year	Sweden
	HussA 1992[h][66]	52	n/a	n/a	n/a	n/a	n/a	1x 22 min session 12 wks	United States
Kirk 2013 [44]	Guendelman 2002 [49]	134	8-16	58	AA 76.1%	128/134	4.5	12 wks	United States
	Jan 2007 [50]	179	6-12	38	n/a	164/179	8.4	12 wks	Taiwan

[a] (i) intervention group; (c) control group

[b] AA=African American.

[c] Stinson provided “participants” and “drop out %” but it is unclear if participants refers to original sample size, or those available for follow-up. Numbers provided were Jan 2007 (164), Joseph 2007 (314), and Krishna 2003 (228).

[d] % provided for individual groups (3-arm trial), however, it stated being unable to provide % for third group due to reporting discrepancies.

[e] Huss 2003: no results were provided in the review for data extraction therefore it was excluded.

[f] There were two trials led by Huss: for the purposes of this review they are referred to as HussA and HussB.

[g] Throughout the Welsh review, this RCT is referred to as “Kamps 2008”, but the reference states 2004.

[h] This review references 3 papers for this trial [66-68], but for the purposes of this review we will reference the most recent publication [66].

Table 3. Descriptions of interventions included within reviews (the checkmark indicates evidence of the presence of an intervention component).

First author of RCT, year	Asthma information, self-care education	Asthma action plan	Self-monitoring, eg, PEF, symptoms with things like diaries	Interactive/ immediate feedback from device	Messages /alerts to patients from device	Message/ alert to/from health professionals	Games, quizzes, vignettes	Daily use	Mode of delivery
Rubin 1986 [60]							√		Computer game
HussA 1992 [66]	√			√					Computer program
Bartholomew 2000 [55]	√		√	√			√	√	CD-ROM
Homer 2000 [56]	√		√	√			√	√	Computer game
Shegog 2001 [57]	√		√	√			√	√	CD-ROM
Vilozni 2001 [62]	√						√		Computer game
Guendelman 2002 [49]	√		√	√		√	√	√	Internet enabled device
HussB 2003 [58]	√			√			√		Computer game
Krishna 2003 [54]	√		√	√		√	√	√	Internet enabled CD ROM
Kamps 2004 [64]	√								Computer program
Shames 2004 [61]	√			√			√		Computer game
Sundberg 2005 [65]	√								Computer program
Rasmussen 2005 [51]		√	√	√	√	√			Web-based
McPherson 2006 [59]				√			√		Computer game
Chan 2007 [63]	√					√			Web-based
Cruz-Correia 2007 [48]	√	√	√	√	√	√			Web-based
Jan 2007 [50]	√	√	√	√	√	√			Web-based
Joseph 2007 [53]	√					√			Web-based
Van der Meer 2009 [52]	√	√	√	√	√	√		√	Web-based

Quality of Evidence in Included Reviews

Seven of the reviews provided risk of bias data based on the guidance provided by the Cochrane collaboration [69], whereby different elements such as adequate sequence generation, allocation concealment, and blinding are assessed as being of low risk of bias, high risk of bias, or unclear risk of bias. This does not provide an overall score. Two reviews provided quality scores, one based on the Oxford Quality scoring system [47] and the other using the Consort Statement [34]. The final review [10] provided an overall quality grade based on guidance from within the Cochrane handbook, but did not provide any rationale for their grading. Grading of RCTs in this way is subjective, and there was conflict between reviews about the risk of bias present in a given RCT. For example, the presence of adequate sequence generation by Guendelman was assessed by 4 different reviews [19,22,42-44]. Two described this as "unclear risk of bias", but the other two reported it as "adequate", and "low risk of bias". This, combined with the various different methods used and in some cases limited information provided, meant that we are unable to make any detailed statements about the quality of the included trials.

Descriptions of Included Randomized Controlled Trials

The 19 unique RCTs within the reviews are described in Table 2. Reporting of descriptive data about RCT participants was mixed; only 4/10 reviews [19-21,44] provided information for all 7 descriptive headings (sample size, age of participants, gender, ethnicity, dropout rate, duration, country of study), with no data about socioeconomic status. The sample sizes for the 18 RCTs providing quantitative results ranged from 20 to 378. There were discrepancies in the sample sizes reported for some RCTs between reviews; for example, the RCT by Jan had 3 different sample sizes reported from 4 reviews (164, 179, and 196). Where there were discrepancies, taking the largest number provided the total number of participants of 2315. From the 19 RCTs featured, 4 were aimed at adults, 14 at children or adolescents, and the age range for one was not described (HussA). All 4 of the RCTs aimed at adults had upper age limits (of 25, 45, 50, and 65 years).

From the 8 RCTs with information about gender, 9 had a majority of female participants and 9 had a majority of males; combining the RCTs with sample size and gender numbers, the percentage of male participants was 54% (793/1478). Dropout rates were available for 11 of the 19 RCTs and ranged from 4.5% to 23%, but again, the reporting of these numbers for 3 of these RCTs was conflicting (Krishna, Jan, Guendelman). Reasons for dropout were rarely provided. Duration of studies ranged from a one-off use of the intervention, to access for 12 months. Eleven RCTs were from the United States, and one each from Taiwan, Portugal, Sweden, Denmark, and the Netherlands. The country was not described for the remaining 3 RCTs.

Descriptions of Contents of Included Interventions

A summary of the key components or ingredients reported as being present in the interventions is summarized in Table 3. Provision of information and self-management education was the most common feature present in 16/19 RCTs. This was followed by the presence of immediate feedback/interactivity from the device (13/19).

Ten of the interventions used games/quizzes/vignettes, and all of these were from RCTs from 2006 or earlier. Eight of the interventions involved some form of direct communication either to or from health professionals. Six of the interventions were available for daily use. The presence of an action plan was noted in 4 interventions, all 2005 or later. This was the same for interventions featuring automated reminders/alerts from devices. Description of the interventions was variable between reviews, and it is possible that many of these interventions feature components not described in Table 3.

The key ingredients or components of interventions were often poorly described rendering it impossible to draw conclusions about the effects of different components of interventions on outcomes.

Results Relating to Outcomes of Interest

There were quantitative results available for the following outcomes of interest: activity limitation, knowledge, markers of self-care, quality of life, medication use, symptoms, missing school, lung function, and health service utilization (eg, emergency department visits, hospitalizations, primary care visits).

There were descriptive results relating to adverse events, behavior change theory use, and patient satisfaction. No data were provided about cost-effectiveness or biomarkers of airway inflammation.

Results From Quantitative Reviews

Quantitative results were provided in a range of ways across the reviews. For example, McLean et al [22] provided original numbers of events and sample size, with results given as odds ratios with confidence intervals, whereas Stinson et al [34] reported the outcome as either statistically significantly in favor of the intervention group (+), the control group (-), or no difference (0). Murray et al [19] presented results as either standardized mean differences (SMD) and effect size (in the form of Lipsey Categories of small, medium, or large) or odds ratios. However, the significance for individual RCT results was not provided, limiting the conclusions we can draw from this individual review (therefore not included in the text below). Quantitative results are presented in Table 4 and Multimedia Appendix 3 and summarized below. Multimedia Appendix 3 details all the data extracted from the reviews, along with a summary stating whether the results favored the intervention group (Y), showed no difference (0), or statistical significance of results was not provided (n/a). No results favored the control groups. Table 4 summarizes only the results that provide a measure of statistical significance [10,20-22,34,37,42-44,47].

Table 4. Table showing statistically significant results (each bullet represents an individual RCT that demonstrates a statistically significant result for a given outcome).

Outcome	Systematic review	RCTs with results showing no difference (sample size)[a,b]	RCTs with results favoring intervention (sample size)[a,b]	Number of RCTs described with a statistically significant result that:		
				Favors control	Shows no difference	Favors intervention
Activity limitation	Johansen [42,43]; Kirk [44]; Stinson [34]		Joseph (314); Guendelman (134)[c]			••
Knowledge	Johansen [42,43]; Kirk [44]; Stinson [34]		Jan (196)[c]; Krishna (228)			••
Markers of self-care	Hieftje [47]; Johansen [42,43]; Kirk [44]; McDermott [37]; Ring [10]; Stinson [34]	Guendelman (134)	Jan (196)[c]; Joseph (314); Rasmussen (300); Rubin (65); Vilonzi (112); HussA (52)		•	••••••
Quality of life	Hieftje [47]; Johansen [42,43],[44]; McLean [22]; Stinson [34]; Welsh [21]	Joseph (314); Krishna (228); Kamps (20)	Jan (196)[c]; Van der Meer (200)[c]; Shames (119); Rasmussen (300)		•••	••••
Medication use	Hieftje [47]; McDermott [37]	Shames (119)	McPherson (101); HussA (52)		•	••
Symptoms	Hieftje [47]; Johansen [42,43]; Kirk [44]; McDermott [37]; McLean [22]; Stinson [34]; Welsh [21]	Guendelman (134); HussA (52); Shames (119); Kamps (20); Sundberg (97); HussB (148)	Jan (196)[c]; Joseph (314); Krishna (228); Rasmussen (300)[c]; Van der Meer (200)[c]; Bartholomew (171)		••••••	••••••
Missing school	Hieftje [47]; McLean [22]; Stinson [34]	Guendelman (134); Rubin (65)	Joseph (314); McPherson(101)		••	••
Lung function	Boyd [20]; Johansen [42,43]; Kirk [44]; McDermott [37]; McLean [22]; Welsh [21]	Jan (196)[c]; Shames (119); Kamps (20); Huss (20)	Rasmussen (300)[c]; Guendelman (134)[c]; Sundberg (97)		••••	•••
Emergency department visits	Hieftje [47]; Johansen [42,43]; Kirk [44]; McLean [22]; Stinson [34]	Rasmussen (300); Bartholomew (171); Jan (196); Guendelman (134)[c]	Joseph (314); Krishna (228)		••••	••
Hospitalization	Hieftje [47]; Johansen [42,43]; Kirk [44]; McLean [22]; Stinson [34]	Guendelman (134)[c]; Joseph (314); Rasmussen (300); Rubin (65); McPherson (101)			•••••	

Outcome	Systematic review	RCTs with results showing no difference (sample size)[a,b]	RCTs with results favoring intervention (sample size)[a,b]	Number of RCTs described with a statistically significant result that:		
				Favors control	Shows no difference	Favors intervention
Primary care visits	Hieftje [47]; Kirk [44]	Shames (119); Rubin (65); McPherson (101); Jan (196)			••••	

[a] References for individual RCTs as per Table 3.

[b] Where there is discrepancy in sample size reporting, the largest sample size is used.

[c] RCT present in more than one systematic review for a given outcome.

Activity Limitation

Stinson [34], Johansen et al [42,43], and Kirk [44] reported findings from two RCTs (Joseph and Guendelman). Both reported that the use of a digital intervention reduced the number of days of restricted activity significantly.

Knowledge

Three reviews [34,42-44] provided results from 2 RCTs (Jan, Krishna) about the impact of online interventions on knowledge, both of which provided results in favor of the use of digital interventions.

Markers for Self-Management

Six of the reviews [10,34,37,42-44,47] presented data for markers of self-management from 7 separate RCTs (Jan, Joseph, Rubin, Vilozni, Guendelman, HussA & Rasmussen) covering, for example, proportion using action plans, spirometry/inhaler technique, diary adherence, and impact on completion of self-management sessions. All except one (Guendelman) showed a positive effect.

Quality of Life

Six reviews [21,22,34,42-44,47] presented data from 7 RCTs for quality of life. Four of the RCTs favored the intervention group (Jan, Van der Meer, Shames, Rasmussen), and 3 showed no difference (Joseph, Krishna, Kamps).

Medication Use

Two reviews [37,47] provided results from 3 RCTs for this outcome. Two favored the interventions group (McPherson & Huss), while the other showed no difference (Shames).

Symptoms and Asthma Control

Seven systematic reviews [21,22,34,37,42-44,47] provided results for this outcome, from 12 RCTs. Six reported no difference (Guendelman, HussA, HussB, Shames, Kamps, Sundberg); the remaining 6 studies favored the use of the digital interventions (Jan, Joseph, Krishna, Rasmussen, Van der Meer, Bartholomew).

Missing School

Three reviews reported for this outcome [22,34,47] from 4 RCTS: two favoring the intervention group (Joseph & McPherson), and two showing no difference (Guendelman & Rubin).

Lung Function

Six reviews [20-22,37,42-44] presented data for this outcome from 7 RCTs, with three favoring the use of digital interventions (Rasmussen, Guendelman, Sundberg), and 4 showing no difference (Jan, Shames, Kamps, HussA).

Emergency Department Visits

Five reviews [22,34,42-44,47] provided results from 6 RCTs on emergency department (ED) visits. Four trials showed no difference (Rasmussen, Bartholomew, Jan, Guendelman), and 2 studies favored the intervention (Joseph, Krishna).

Hospitalization

Five reviews [22,34,42-44,47] provided information from 5 RCTs, all of which showed no significant differences between the intervention group and controls (Guendelman, Joseph, Rubin, McPherson, Rasmussen).

Primary Care Visits

Two reviews provided data for this outcome [44,47] from 4 RCTs (Shames, Rubin, McPherson & Jan). All 4 RCTs demonstrated no difference in the number of visits.

Outcomes With Descriptive Results Only

Adverse Events

Only two reviews [22,37] provided results for this outcome. McLean et al [22] results were specific to one study (Rasmussen) and found that the increased corticosteroid dose/use that went along with being in the intervention or specialist group, compared to the GP group, meant a higher proportion (no details provided) of those patients experienced dysphonia or oral candidiasis. McDermott made comments more generally stating that there appeared to be no adverse outcomes from moving towards computer-based patient self-management programs.

Use of Behavior Change Theory During Development

Two systematic reviews explicitly sought to establish the presence or absence of underlying theory in the development of their included interventions [42-44]. Johansen et al [42,43] included as part of their quality appraisal whether or not there was “theoretical evidence that the intervention might have the desired effect”. Of the 5 digital asthma interventions included in their review, all 5 met this criterion (Chan, Guendelman, Jan, Rasmussen, Van der Meer). Kirk et al [44] described the presence or absence of an “underlying theoretical basis” in their included RCTs. They reported that both the digital asthma RCTs in their review (Jan and Guendelman) demonstrated no theoretical basis, which contradicts the findings of Johansen et al [42,43].

Three other reviews described the contents of their included interventions using terminology relevant to behavior change theory. For example, within Murray et al [19], the Shegog trial was described as providing "intensive, tailored information on self-management for children with asthma. Text, graphics, animation, sound and video clips are utilised, and behavior support delivered via verbal reinforcement, guided practice, feedback goal setting and incentives". However, it does not explicitly describe any role of behavior change theory during development of the intervention.

Stinson et al [34] reported that all three of their included interventions (Jan, Joseph, Krishna) featured training around symptom management, trigger avoidance, and medication use, but that Krishna was the only one described as including behavioral therapy (featuring modeling, reinforcement, and self-mastery), but again no explicit descriptions of theory used during development.

McDermott et al [37] did not describe the presence or absence of behavior change theories during the development of interventions included in their review but did attempt to characterize the "active ingredients" of each included intervention by coding their descriptions using a behavior change technique (BCT) taxonomy. One intervention (HussA) was found to include only one BCT (provide instruction) while the other (Sundberg) was found to have 3 BCTs (provide general information on the condition, advise on medication, and provide instruction).

Patient Satisfaction, Barriers, and Facilitators to Digital Asthma Interventions

Three reviews [22,37,42,43] addressed these issues. One study (Cruz-Correia) featured within the McLean et al [22] review found that patients preferred a Web-based system of monitoring asthma compared to a paper-based system. Johansen et al [42,43] reported in general that "an interactive-feedback learning mechanism can provide the stimulus for the patient to build the necessary confidence to handle symptoms and self-management, and in this way support patient centeredness", and they reported specifically that the Guendelman study suggested the electronic devices might be considered as a "motivating and exciting tool for children with asthma". Johansen et al [42,43] also reported the findings from the trial by Jan that children found their tool to be "fun" and concluded more research was required in this area. McDermott et al [37] reported findings in general suggesting that computer-based programs may "even be preferred by many patients as they allow participants to proceed at their own pace". McDermott et al [37] also commented that the combination of standard and computer-based approaches might seem to be the ideal scenario, but state there was no evidence found in their review that is the case.

Only one review mentioned implementability of interventions. Kirk et al [44] felt that the integration of the intervention featured in the trial led by Jan was feasible in current practice but did not elaborate.

Cost Effectiveness Analysis

Three Cochrane reviews [19,20,22] and one narrative synthesis planned to include this outcome measure. McLean et al [22] looked at "costs from the health care perspective", but there were no results specific to interventions matching our inclusion criteria. Murray et al [19] looked at "economic outcomes", including health care use under this heading. Other than providing summaries of health care use for 3 trials (Bartholomew, Guendelman, Krishna), there was no additional specific cost data provided. Boyd et al [20] planned to collect data on cost, but no studies provided such information. McDermott et al [37] found only one of their included studies provided data on cost, and it was for an intervention aimed at those with diabetes.

Discussion

Principal Findings

Initial full paper screening resulted in 29 systematic reviews. However, following quality appraisal, only 10 met the predetermined AMSTAR cut-off values. This metareview summarizes the findings from these 10 systematic reviews featuring a total of 18 unique RCTs of digital asthma self-management interventions. Only 11 of the 19 RCTs were present in more than one review suggesting that at present there is no single good quality systematic review looking specifically at digital self-management interventions for asthma.

Within the systematic reviews we found that information about methods, intervention components, and results were often brief and did not allow for meaningful comparisons between interventions. The described studies themselves were extremely heterogeneous making comparison between interventions difficult. "Control" groups ranged from no active intervention or contact with health professionals to multiple face-to-face teaching sessions and intermittent use of the intervention itself, which may have masked potential positive outcomes of digital interventions. More recent reviews tended to have more information, particularly with online appendices.

In no studies did the control groups have better outcomes, although only two reviews addressed the issue of adverse events, and specific information was available from only one study, which is a concern. That study suggested digital interventions groups may be at higher risk of adverse events related to the fact that successful interventions usually increase inhaled corticosteroid use (a positive outcome) and therefore results in more cases of dysphonia or oral candidiasis. Surprisingly, the issue of patient satisfaction was also neglected, being specifically addressed in only one trial, which suggested participants preferred a Web-based system to a paper-based one, and discussed in general terms in one further review.

Descriptions of intervention development and particularly the use of theory (which has been shown to increase effectiveness) were either brief or not discussed at all, although our review suggests that more recent reviews are increasingly recognizing this, an issue highlighted by the recent publication of a CONSORT EHEALTH statement [70]. This matters because trying to establish the key components of effective self-management interventions is a challenge faced by researchers in this area [71]. For example, a recent Cochrane review of computer-based interventions in adults with type 2 diabetes concluded that there was a small beneficial effect on blood glucose control, particularly in the mobile phone subgroup [72]. However, the authors also commented that the key “ingredient” of effective interventions was unclear; it could either have been that the mobile phone itself was important or that the interventions delivered using mobile phones had included BCTs that were likely to lead to success, and these could be the “ingredient” that made these interventions more effective. What is consistent across several systematic reviews on digital support for self-management of chronic illness is that interventions with multiple behavior change techniques appear, on the whole, to be more effective than those using fewer and that the use of theory to inform the choice and combination of BCTs appears to be associated with increasing effectiveness [72-74].

Only four RCTs featured adults, and there were no participants over the age of 65 years, with only one small trial with just 21 participants including individuals over the age of 50 years. No information was provided about the socioeconomic status or education level of participants. None of the trials were undertaken in low-income countries. Attrition rates of up to 23% were recorded. This is in keeping with other systematic reviews of digital interventions in other areas. For example, a recent Cochrane review examining computer-based weight loss interventions found attrition rates ranging from 2-25% (median 16%) [75], and attrition may be worse in interventions targeting older age groups, with one review including digital and nondigital interventions noting rates between 0 and 52% (median 15%) [73]. Reassuringly, attrition rates are no worse than those found with nondigital self-management asthma interventions as described in Gibson’s Cochrane review examining asthma self-management education and regular health professional review, where attrition rates ranged from 0-54% (median 15%) [8].

Our metareview suggests that digital interventions may be effective at improving knowledge, reducing activity limitation, improving markers of self-management, improving quality of life, and optimizing medication use in those less than 65 years of age. However, certain indicators, such as knowledge and activity limitation, were assessed in only two trials, and medication use improvements noted in two out of three trials that examined this. There was no evidence of improvements in symptoms, lung function, school absences, or health service utilization.

Importantly, we found no qualitative synthesis of asthma digital interventions that would have given insight into the patient experience of using digital self-management interventions, implementation processes, or barriers and facilitators to their use, which is an important gap in the evidence base.

Strengths and Limitations

Our metareview has a number of strengths and limitations. The search was undertaken by a team with good experience of systematic reviews, using multiple databases, and using a strategy designed iteratively with researchers to be as inclusive as possible, without being unwieldy. Despite this, we may have missed reviews of chronic illness interventions including asthma but not specifically indexed as such. We included non-English studies, which is a strength. Due to the heterogeneity of the data, a formal metasynthesis was not possible.

A further limitation here is the reliability and comprehensibility of the included information. When undertaking a review of reviews, the data are an extra step away from the original research increasing the possibility of reporting errors. We noted several discrepancies between reviews describing the same RCTs, in sample sizes and gender descriptions. Some trials presented data as sample size, while others used terms such as number of participants, and dropout calculations did not appear to correlate on several occasions. This could be due to reporting error, or differences in the interpretation of the terms sample size, follow-up, and participants. In addition, there was a lack of detail describing the populations included, and limited definitions of outcome measures provided by featured systematic reviews. This translates into a lack of specific detail about the population our results are relevant to, and limitations in comparing results that have potentially used different ways of defining outcomes. Finally, establishing a minimum standard of quality of included reviews ensures that there is a degree of reliability to the conclusions [76]; however, this undoubtedly narrowed the available data from which we could draw conclusions. Had we not used AMSTAR criteria, we could have had data from up to a further 12 RCTs and two qualitative studies. However many of these RCTs were included only in low AMSTAR scoring papers (ie, those that had not undertaken any quality appraisal of their included studies), and there were often no descriptions of control groups, or more than a few words describing the interventions, and therefore inclusion of any such data would have rendered our conclusions meaningless. The RCTs themselves often appeared poor quality (eg, poor randomization strategies), with small numbers (eg, sample sizes as low as 10), and had been excluded from the more robust reviews for these reasons. The lack of economic data is a weakness, although the results on health care resource use (hospitalizations and ED visits) suggests that evidence of cost-effectiveness may be lacking. However, without data including routine health care resource utilization and formal economic analysis, no firm conclusions can be drawn.

Conclusions

This metareview provides a snapshot of current knowledge about effectiveness of digital self-management support for those with asthma. Digital self-management interventions show promise, with evidence of beneficial effects on some outcomes. However, we know nothing about the socioeconomic status of participants, and few over the age of 50 years and no one over 65 years of age were included. Thus, the true “reach” of these studies is uncertain, and their likely uptake and use by the wider population of those with asthma remains uncertain. Few interventions were underpinned by robust theoretical frameworks. Digital interventions are poorly described, and there is insufficient information about barriers and facilitators to their uptake and utilization. Importantly, patient perspectives have been largely ignored in currently available reviews. There was little data about cost effectiveness within reviews, but this appears to relate to the lack of existence of such data from primary trials. Digital interventions for asthma appear promising but further robust investigation is needed, first, in the form of a detailed systematic review of currently available digital interventions aimed at those with asthma, detailing the presence or absence of BCTs. Second, examination of the primary qualitative literature to describe what is already known about the patient's perspective would be invaluable to inform future interventions.

Acknowledgments

We would like to acknowledge the York Health Economic Consortium for their work on the search strategy. We would like to thank the Chief Scientist Office Scotland for funding this study (Ref: CAF/11/08). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

Authors met the following criteria for author contributions: substantial contribution to conception and design, or acquisition of data, or analysis and interpretation of data (DM, SW, KA, EC, RD, AMcC, AMM, VR, NCT, FSM); drafting the article or revising it critically for important intellectual content (DM, SW, KA, EC, AMcC, NCT, FSM); final approval of the version to be published (DM, SW, KA, EC, RD, AMcC, AMM, VR, NCT, FSM).

Conflicts of Interest

DM, KA, EC, RD, AMcC, AMM, VR, NCT, and FSM have no conflict of interest. SW is co-author on one included review (Ring et al) but took no part in the screening, selection, or quality appraisal of this study.

Multimedia Appendix 1

Full systematic review protocol.

[PDF File (Adobe PDF File), 29KB - jmir_v16i2e51_app1.pdf]

Multimedia Appendix 2

Details and results of databases searched, and sample search strategy for Medline Database.

[PDF File (Adobe PDF File), 41KB - jmir_v16i2e51_app2.pdf]

Multimedia Appendix 3

Description of full quantitative results per outcome.

[PDF File (Adobe PDF File), 76KB - jmir_v16i2e51_app3.pdf]

References

1. Masoli M, Fabian D, Holt S, Beasley R, Global Initiative for Asthma (GINA) Program. The global burden of asthma: executive summary of the GINA Dissemination Committee report. Allergy 2004 May;59(5):469-478. [doi: 10.1111/j.1398-9995.2004.00526.x] [Medline: 15080825]
2. Global Strategy for Asthma Management and Prevention. 2012. Global Initiative for Asthma (GINA) URL: http://www.ginasthma.org/local/uploads/files/GINA_Report_March13.pdf [accessed 2013-06-25] [WebCite Cache ID 6HdiIhxv6]
3. Rabe KF, Vermeire PA, Soriano JB, Maier WC. Clinical management of asthma in 1999: the Asthma Insights and Reality in Europe (AIRE) study. Eur Respir J 2000 Nov;16(5):802-807 [FREE Full text] [Medline: 11153575]
4. Partridge MR, Dal Negro RW, Olivieri D. Understanding patients with asthma and COPD: insights from a European study. Prim Care Respir J 2011 Sep;20(3):315-23, 17 p following 323 [FREE Full text] [doi: 10.4104/pcrj.2011.00056] [Medline: 21660394]
5. British Thoracic Society Scottish Intercollegiate Guidelines Network. British Guideline on the Management of Asthma. Thorax 2008 May;63 Suppl 4:iv1-121. [doi: 10.1136/thx.2008.097741] [Medline: 18463203]
6. Powell H, Gibson PG. Options for self-management education for adults with asthma. Cochrane Database Syst Rev 2003(1):CD004107. [doi: 10.1002/14651858.CD004107] [Medline: 12535511]
7. Thoonen BP, Schermer TR, Van Den Boom G, Molema J, Folgering H, Akkermans RP, et al. Self-management of asthma in general practice, asthma control and quality of life: a randomised controlled trial. Thorax 2003 Jan;58(1):30-36 [FREE Full text] [Medline: 12511716]

8. 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]
9. Hoskins G, McCowan C, Donnan PT, Friend JA, Osman LM, Asthma UK Scotland, Scottish Executive Health Department. Results of a national asthma campaign survey of primary care in Scotland. Int J Qual Health Care 2005 Jun;17(3):209-215 [FREE Full text] [doi: 10.1093/intqhc/mzi036] [Medline: 15831548]
10. Ring N, Malcolm C, Wyke S, Macgillivray S, Dixon D, Hoskins G, et al. Promoting the use of Personal Asthma Action Plans: a systematic review. Prim Care Respir J 2007 Oct;16(5):271-283 [FREE Full text] [doi: 10.3132/pcrj.2007.00049] [Medline: 17710351]
11. 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]
12. Eccles MP, Armstrong D, Baker R, Cleary K, Davies H, Davies S, et al. An implementation research agenda. Implement Sci 2009;4:18 [FREE Full text] [doi: 10.1186/1748-5908-4-18] [Medline: 19351400]
13. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
14. Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012 May 1;90(5):357-364 [FREE Full text] [doi: 10.2471/BLT.11.099424] [Medline: 22589569]
15. O'Connor D, Green S, Higgins JPT. Chapter 5: Defining the review question and developing criteria for including studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester, England: Wiley-Blackwell; 2008.
16. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009 Oct;62(10):1013-1020. [doi: 10.1016/j.jclinepi.2008.10.009] [Medline: 19230606]
17. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]
18. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One 2007;2(12):e1350 [FREE Full text] [doi: 10.1371/journal.pone.0001350] [Medline: 18159233]
19. 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]
20. Boyd M, Lasserson TJ, McKean MC, Gibson PG, Ducharme FM, Haby M. Interventions for educating children who are at risk of asthma-related emergency department attendance. Cochrane Database Syst Rev 2009(2):CD001290. [doi: 10.1002/14651858.CD001290.pub2] [Medline: 19370563]
21. Welsh EJ, Hasan M, Li P. Home-based educational interventions for children with asthma. Cochrane Database Syst Rev 2011(10):CD008469. [doi: 10.1002/14651858.CD008469.pub2] [Medline: 21975783]
22. McLean S, Chandler D, Nurmatov U, Liu J, Pagliari C, Car J, et al. Telehealthcare for asthma. Cochrane Database Syst Rev 2010(10):CD007717. [doi: 10.1002/14651858.CD007717.pub2] [Medline: 20927763]
23. 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]
24. Cushing CC, Steele RG. A meta-analytic review of eHealth interventions for pediatric health promoting and maintaining behaviors. J Pediatr Psychol 2010 Oct;35(9):937-949 [FREE Full text] [doi: 10.1093/jpepsy/jsq023] [Medline: 20392790]
25. Coffman JM, Cabana MD, Halpin HA, Yelin EH. Effects of asthma education on children's use of acute care services: a meta-analysis. Pediatrics 2008 Mar;121(3):575-586 [FREE Full text] [doi: 10.1542/peds.2007-0113] [Medline: 18310208]
26. Krishna S, Balas EA, Spencer DC, Griffin JZ, Boren SA. Clinical trials of interactive computerized patient education: implications for family practice. J Fam Pract 1997 Jul;45(1):25-33. [Medline: 9228910]
27. Lewis D. Computers in patient education. Comput Inform Nurs 2003;21(2):88-96. [Medline: 12802949]
28. Almeida P, Caetan F, Carvalho A, Fernandes S, Gaspar J, Loureiro M, et al. Introduction to Medicine. 2006. Interactive Applications for patient education and behaviour change in Asthma URL: http://medicina.med.up.pt/im/trabalhos06_07/sites/Turma16/sitee/articledownload.pdf [accessed 2013-07-02] [WebCite Cache ID 6Ho8BrW1f]
29. Bussey-Smith KL, Rossen RD. A systematic review of randomized control trials evaluating the effectiveness of interactive computerized asthma patient education programs. Ann Allergy Asthma Immunol 2007 Jun;98(6):507-16; quiz 516, 566. [doi: 10.1016/S1081-1206(10)60727-2] [Medline: 17601262]
30. García-Lizana F, Sarría-Santamera A. New technologies for chronic disease management and control: a systematic review. J Telemed Telecare 2007;13(2):62-68. [doi: 10.1258/135763307780096140] [Medline: 17359568]

31. Fox MP. A systematic review of the literature reporting on studies that examined the impact of interactive, computer-based patient education programs. Patient Educ Couns 2009 Oct;77(1):6-13. [doi: 10.1016/j.pec.2009.02.011] [Medline: 19345550]
32. Moeinedin F, Moineddin R, Jadad AR, Hamid JS, To T, Beyene J. Application of biomedical informatics to chronic pediatric diseases: a systematic review. BMC Med Inform Decis Mak 2009;9:22 [FREE Full text] [doi: 10.1186/1472-6947-9-22] [Medline: 19416540]
33. Papastergiou M. Exploring the potential of computer and video games for health and physical education: A literature review. Computers & Education 2009 Nov;53(3):603-622. [doi: 10.1016/j.compedu.2009.04.001]
34. Stinson J, Wilson R, Gill N, Yamada J, Holt J. A systematic review of internet-based self-management interventions for youth with health conditions. J Pediatr Psychol 2009 Jun;34(5):495-510 [FREE Full text] [doi: 10.1093/jpepsy/jsn115] [Medline: 19029142]
35. Coscrato G, Pina JC, Mello DFD. Utilização de atividades lúdicas na educação em saúde: uma revisão integrativa da literatura [Use of recreational activities in health education: integrative review of literature]. Acta paul. enferm 2010 Apr;23(2):257-263. [doi: 10.1590/S0103-21002010000200017]
36. Gremeaux V, Coudeyre E. The Internet and the therapeutic education of patients: A systematic review of the literature. Ann Phys Rehabil Med 2010 Dec;53(10):669-692. [doi: 10.1016/j.rehab.2010.09.003] [Medline: 21036689]
37. McDermott MS, While AE. Maximizing the healthcare environment: a systematic review exploring the potential of computer technology to promote self-management of chronic illness in healthcare settings. Patient Educ Couns 2013 Jul;92(1):13-22. [doi: 10.1016/j.pec.2013.02.014] [Medline: 23566427]
38. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010;12(2):e21 [FREE Full text] [doi: 10.2196/jmir.1357] [Medline: 20554500]
39. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
40. Kuo C, Wang H. Applying self-management intervention studies to patients with chronic illness: a systematic review [Chinese]. Journal of Nursing and Healthcare Research 2012(8):50-60 ISSN 2072-90235.
41. Chrisler AJ. Asthma education: a review of randomized controlled youth asthma-education programs. J Public Health 2012 Oct 27;20(3):305-312. [doi: 10.1007/s10389-011-0449-8]
42. Johansen MA, Berntsen GK, Schuster T, Henriksen E, Horsch A. Electronic symptom reporting between patient and provider for improved health care service quality: a systematic review of randomized controlled trials. Part 2: methodological quality and effects. J Med Internet Res 2012;14(5):e126 [FREE Full text] [doi: 10.2196/jmir.2216] [Medline: 23032363]
43. Johansen MA, Henriksen E, Horsch A, Schuster T, Berntsen GK. Electronic symptom reporting between patient and provider for improved health care service quality: a systematic review of randomized controlled trials. Part 1: state of the art. J Med Internet Res 2012;14(5):e118 [FREE Full text] [doi: 10.2196/jmir.2214] [Medline: 23032300]
44. Kirk S, Beatty S, Callery P, Gellatly J, Milnes L, Pryjmachuk S. The effectiveness of self-care support interventions for children and young people with long-term conditions: a systematic review. Child Care Health Dev 2013 May;39(3):305-324. [doi: 10.1111/j.1365-2214.2012.01395.x] [Medline: 22676438]
45. Mosnaim GS, Powell LH, Rathkopf M. A review of published studies using interactive Internet tools or mobile devices to improve asthma knowledge or health outcomes. Pediatric Allergy, Immunology, and Pulmonology 2012 Jun;25(2):55-63. [doi: 10.1089/ped.2011.0112]
46. Nickels A, Dimov V. Innovations in technology: social media and mobile technology in the care of adolescents with asthma. Curr Allergy Asthma Rep 2012 Dec;12(6):607-612. [doi: 10.1007/s11882-012-0299-7] [Medline: 22976493]
47. Hieftje K, Edelman EJ, Camenga DR, Fiellin LE. Electronic media-based health interventions promoting behavior change in youth: a systematic review. JAMA Pediatr 2013 Jun;167(6):574-580 [FREE Full text] [doi: 10.1001/jamapediatrics.2013.1095] [Medline: 23568703]
48. Cruz-Correia R, Fonseca J, Lima L, Araújo L, Delgado L, Castel-Branco MG, et al. Web-based or paper-based self-management tools for asthma--patients' opinions and quality of data in a randomized crossover study. Stud Health Technol Inform 2007;127:178-189. [Medline: 17901611]
49. 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. [Medline: 11814370]
50. 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]
51. Rasmussen LM, Phanareth K, Nolte H, Backer V. Internet-based monitoring of asthma: a long-term, randomized clinical study of 300 asthmatic subjects. J Allergy Clin Immunol 2005 Jun;115(6):1137-1142. [doi: 10.1016/j.jaci.2005.03.030] [Medline: 15940125]
52. 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]

53. 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]
54. Krishna S, Francisco BD, Balas EA, König P, Graff GR, Madsen RW, Randomized trial. Internet-enabled interactive multimedia asthma education program: a randomized trial. Pediatrics 2003 Mar;111(3):503-510. [Medline: 12612228]
55. Bartholomew LK, Gold RS, Parcel GS, Czyzewski DI, Sockrider MM, Fernandez M, et al. Watch, Discover, Think, and Act: evaluation of computer-assisted instruction to improve asthma self-management in inner-city children. Patient Educ Couns 2000 Feb;39(2-3):269-280. [Medline: 11040726]
56. Homer C, Susskind O, Alpert HR, Owusu M, Schneider L, Rappaport LA, et al. An evaluation of an innovative multimedia educational software program for asthma management: report of a randomized, controlled trial. Pediatrics 2000 Jul;106(1 Pt 2):210-215. [Medline: 10888694]
57. Shegog R, Bartholomew LK, Parcel GS, Sockrider MM, Mâsse L, Abramson SL. Impact of a computer-assisted education program on factors related to asthma self-management behavior. J Am Med Inform Assoc 2001;8(1):49-61 [FREE Full text] [Medline: 11141512]
58. Huss K, Winkelstein M, Nanda J, Naumann PL, Sloand ED, Huss RW. Computer game for inner-city children does not improve asthma outcomes. J Pediatr Health Care 2003;17(2):72-78. [doi: 10.1067/mph.2003.28] [Medline: 12665729]
59. McPherson AC, Glazebrook C, Forster D, James C, Smyth A. A randomized, controlled trial of an interactive educational computer package for children with asthma. Pediatrics 2006 Apr;117(4):1046-1054 [FREE Full text] [doi: 10.1542/peds.2005-0666] [Medline: 16585298]
60. Rubin DH, Leventhal JM, Sadock RT, Letovsky E, Schottland P, Clemente I, et al. Educational intervention by computer in childhood asthma: a randomized clinical trial testing the use of a new teaching intervention in childhood asthma. Pediatrics 1986 Jan;77(1):1-10. [Medline: 3510016]
61. Shames RS, Sharek P, Mayer M, Robinson TN, Hoyte EG, Gonzalez-Hensley F, et al. Effectiveness of a multicomponent self-management program in at-risk, school-aged children with asthma. Ann Allergy Asthma Immunol 2004 Jun;92(6):611-618. [doi: 10.1016/S1081-1206(10)61426-3] [Medline: 15237762]
62. Vilozni D, Barker M, Jellouschek H, Heimann G, Blau H. An interactive computer-animated system (SpiroGame) facilitates spirometry in preschool children. Am J Respir Crit Care Med 2001 Dec 15;164(12):2200-2205. [doi: 10.1164/ajrccm.164.12.2101002] [Medline: 11751188]
63. 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]
64. Kamps AW. Unscheduled care for people with asthma in a multi-ethnic area is reduced following educational outreach programme by specialist nurses. Evidence-based Healthcare 2004 Aug;8(4):190-191. [doi: 10.1016/j.ehbc.2004.05.017]
65. Sundberg R, Tunsäter A, Palmqvist M, Ellbjär S, Löwhagen O, Torén K. A randomized controlled study of a computerized limited education program among young adults with asthma. Respir Med 2005 Mar;99(3):321-328. [doi: 10.1016/j.rmed.2004.08.006] [Medline: 15733508]
66. Huss K, Huss RW, Squire EN, Carpenter GB, Smith LJ, Salata K, et al. Computer education for asthmatics: what effects? J Nurs Care Qual 1992 Apr;6(3):57-66. [Medline: 1550949]
67. Huss K, Salerno M, Huss RW. Computer-assisted reinforcement of instruction: Effects on adherence in adult atopic asthmatics. Res Nurs Health 1991;14(4):259-267. [doi: 10.1002/nur.4770140404]
68. Huss K, Squire E, Carpenter GB, Smith LJ, Huss RW, Salata K, et al. Effective education of adults with asthma who are allergic to dust mites. J Allergy Clin Immunol 1992 Apr;89(4):836-843. [Medline: 1560167]
69. Higgins J, Altman D. Chapter 8: Assessing risk of bias in included studies. In: Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester, UK: John Wiley & Sons; 2008.
70. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
71. Michie S, Fixsen D, Grimshaw JM, Eccles MP. Specifying and reporting complex behaviour change interventions: the need for a scientific method. Implement Sci 2009;4:40 [FREE Full text] [doi: 10.1186/1748-5908-4-40] [Medline: 19607700]
72. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013;3:CD008776. [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]
73. Aalbers T, Baars MA, Rikkert MG. Characteristics of effective Internet-mediated interventions to change lifestyle in people aged 50 and older: a systematic review. Ageing Res Rev 2011 Sep;10(4):487-497. [doi: 10.1016/j.arr.2011.05.001] [Medline: 21628005]
74. 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]

75. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012;8:CD007675. [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
76. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 2011;11(1):15 [FREE Full text] [doi: 10.1186/1471-2288-11-15] [Medline: 21291558]

Abbreviations

AMSTAR: A Measurement Tool to Assess Systematic Reviews
ACQ: Asthma Control Questionnaire
AQLQ: Asthma Quality of Life Questionnaire
BCT: behavior change techniques
ED: emergency department
FEV1: forced expiratory volume in 1 second
OR: odds ratio
PEF: peak expiratory flow rate
PICOS: participants, interventions, comparison, outcomes, study design
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by G Eysenbach; submitted 02.07.13; peer-reviewed by J Shepard, M Belisario; comments to author 18.09.13; revised version received 18.11.13; accepted 12.12.13; published 18.02.14.

Please cite as:
Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, MacKenzie AM, McConnachie A, Raghuvir V, Thomson NC, Mair FS
Digital Asthma Self-Management Interventions: A Systematic Review
J Med Internet Res 2014;16(2):e51
URL: http://www.jmir.org/2014/2/e51/
doi:10.2196/jmir.2814
PMID:24550161

Review

Self-Directed Interventions to Promote Weight Loss: A Systematic Review of Reviews

Jason Tang[1], MSc; Charles Abraham[1], PhD; Colin Greaves[1], PhD; Tom Yates[2,3], PhD

[1]University of Exeter Medical School, University of Exeter, Exeter, United Kingdom

[2]NIHR Leicester-Loughborough Diet, Lifestyle, and Physical Activity Biomedical Research Unit, Leicester, United Kingdom

[3]Diabetes Research Centre, College of Medicine, Biological Sciences and Psychology, University of Leicester, Leicester, United Kingdom

Corresponding Author:
Charles Abraham, PhD
University of Exeter Medical School
University of Exeter
St Luke's Campus, Heavitree Road
Exeter, EX1 2LU
United Kingdom
Phone: 44 07969330607
Fax: 44 01392264872
Email: C.Abraham@exeter.ac.uk

Related Article:

Letter: http://www.jmir.org/2014/7/e178/

Abstract

Background: A wide range of self-directed weight-loss interventions are available, providing users with a variety of tools delivered through various formats to regulate weight-related behavior patterns. However, it is unclear how effective self-directed interventions are and how they promote weight loss and weight maintenance.

Objective: A systematic review of reviews was conducted to examine the effectiveness of such interventions and to identify intervention content associated with effectiveness.

Methods: MEDLINE, Embase, PsycINFO, CINAHL, and the Cochrane Library for systematic reviews were searched from 2000-2012 for reviews of the effectiveness of self-directed interventions on weight loss and weight maintenance in adults. Two reviewers used predefined inclusion criteria to select relevant reviews and assess their quality using the Overview Quality Assessment Questionnaire (OQAQ). We extracted data on effectiveness and on relationships between intervention characteristics and effectiveness.

Results: Twenty reviews were included and quality assessed. Findings relevant to self-directed interventions, including interactive websites, smartphone applications, and text messaging (short message service, SMS) were summarized. Findings were mixed but promising. For example, one review of Internet-based interventions found that, when used in conjunction with standard weight loss programs, these interventions resulted in a significant average increase in weight loss of 1.5 kg over evaluation periods. Unfortunately, only 7 of 20 reviews were of high methodological quality according to OQAQ scores, and only 4 employed meta-analyses. Few reviews linked intervention content to effectiveness.

Conclusions: Current evidence suggests that self-directed interventions can independently promote weight loss and can augment interventions involving personal contact. Particular change techniques and delivery modes including individualized feedback, email counseling, and online social support appear to enhance effectiveness. Further reviews of the content of self-directed weight-loss intervention studies are needed to clarify which change techniques delivered through which delivery formats optimize intervention effectiveness.

(J Med Internet Res 2014;16(2):e58) doi:10.2196/jmir.2857

KEYWORDS

weight loss; obesity; Internet; eHealth; home-based; text message; self-delivered; intervention; systematic review

Introduction

Weight reduction is a global health priority because being overweight or obese is associated with multiple health problems, including the leading causes of preventable death such as cardiovascular disease, type 2 diabetes, and particular cancers [1,2]. Yet the prevalence of obesity and health services resources devoted to treating its consequences are increasing internationally. In the United States, for example, 68% of adults are overweight or obese accounting for more than 20% of health care costs [3].

Pharmaceutical and bariatric surgery treatments are effective for some overweight and obese people but are expensive and often accompanied by adverse side effects. Consequently, they tend to be weight loss treatments of last resort [4]. Reversing population obesity trends depends on prompting widespread changes in diet and physical activity patterns [5,6]. Promotion of initiation and maintenance of weight changes will be optimized by an understanding of the processes regulating eating and physical activity patterns [7,8].

Effective face-to-face interventions have been developed to promote weight loss through changes in diet and physical activity [9], but these require substantial, specialist delivery personnel and resources [10]. Consequently, more intensive (higher contact frequency) and expensive interventions are most effective. Further research is warranted on intervention formats that could reduce costs without decreasing effectiveness [11]. Effective, high-intensity, low-cost interventions may be developed if participants self-deliver intervention content using printed media (eg, self-help manuals) or interactive software (on mobile phones, the Internet, or other online mobile devices). We use the term “self-directed interventions” to mean those that require minimal professional contact (for example, provision of initial instructions) or no professional contact and can be easily used with existing infrastructure and in the context of users’ everyday lives. Many such interventions have been developed [12], and although attrition rates are often high [13], such interventions have been found to be effective for a broad range of health behavior changes including improving diabetes self-management and smoking cessation [14,15].

Self-directed interventions are likely to be most effective when they empower participants to control and regulate their own thoughts, feelings, and behaviors, thereby changing psychological and environmental prompts to weight-gain behaviors [8]. These interventions are not only self-directed but also “self-regulatory” in that people are taught to change the regulatory processes that maintain current behavior patterns and establish new ones. For example, it has been suggested that prompting self-regulation through self-monitoring of behavior, providing timely feedback on behavior changes, prompting goal setting, and specific action planning are all associated with effectiveness in dietary and physical activity interventions [16,11]. Nonetheless, it remains unclear which self-directed weight loss interventions are effective and why. For example, which combinations of behavior change techniques [17,18] targeting which behavior regulation processes delivered through which particular delivery formats [19] optimize weight loss and weight maintenance over time?

Considerable research has been devoted to developing and evaluating self-directed, weight loss interventions, and a number of recent reviews are available. Some reviews have focused solely on studies evaluating interventions using weight loss outcomes [20], while others have included studies evaluating interventions in terms of weight loss alongside studies using other outcome measures such as self-report behavioral measures.

In a systematic review of reviews, Kohl and Crutzen examined the efficacy, use, and reach of Internet-based interventions for lifestyle changes in physical activity, dietary behaviors, smoking, alcohol consumption, and condom use [21]. One meta-analysis included in this review found that Internet-based interventions of longer duration, based on social cognitive theories, and including educational components with regular updates of intervention content increased physical activity levels [22]. These reviewers also reported that interactive elements, such as chat rooms and online peer support, were associated with greater efficacy. However, identification of such components across interventions was rare.

We are not aware of any previous review of reviews of self-directed interventions evaluated in terms of weight loss outcomes. We therefore conducted a systematic review of reviews to summarize efficacy evidence and design features of self-directed interventions designed to reduce weight and sustain weight maintenance. Within identified reviews, we focused on the conclusions that reviewers drew about interventions evaluated in terms of weight loss. This meant that, for some reviews, all the included primary studies were relevant to our research questions, while for others, a minority of the primary studies were relevant.

Our review aimed to summarize evidence in relation to three key questions:

1. How effective are self-directed weight loss interventions?
2. Is effectiveness enhanced by use of particular change techniques?
3. Is effectiveness enhanced by using particular delivery formats?

Methods

Review Inclusion Criteria

To meet these aims, we included reviews based on systematic literature searches published in English between 2000 and 2012 that included at least one primary intervention evaluation:

1. Of an individual-level, self-directed weight loss intervention targeting healthy adults (18 years or over) who were normal weight, sedentary, overweight, or obese. Normal weight intervention participants were included because such studies are important to understanding what works best in prevention of weight gain and maintenance of normal weight in nonclinical populations.
2. Targeting physical activity, diet, or both and were evaluated using at least one weight-related outcome (eg, weight, body mass index [BMI], waist circumference, waist to hip ratio).

3. Employed randomized controlled trials (RCTs), observational, quasi-experimental, and/or cohort studies. Comparison groups could include usual care, other interventions, or no intervention.

Search Strategy

Reviews that met these inclusion criteria were searched for on the bibliographic electronic databases MEDLINE (Ovid), Embase (Ovid), PsycINFO (Ovid), CINAHL, and the Cochrane Library. Full searches applied in each database are available from the authors.

Study Selection

The first author examined the titles and abstracts of articles identified by our search against the predefined inclusion criteria. A second researcher repeated this process, and discrepancies were resolved through discussion. Full text articles were obtained and assessed to ensure correspondence to inclusion criteria by the first 3 authors. Disagreements were resolved through discussion, and reasons for exclusion were outlined for each review. See Multimedia Appendix 1 for a list of included and excluded reviews (n=32). References in eligible reviews were checked to identify further relevant reviews.

Quality Assessment

The quality of each full-text article that met the inclusion criteria was rated by the first and second authors using the Overview Quality Assessment Questionnaire (OQAQ) [23]. Each review was scored against a checklist of nine standard items, including transparency, selection bias, study quality, and replicability. The few scoring disagreements arising were resolved through discussion. Following Greaves et al's review of reviews on components associated with effectiveness in dietary and physical activity intervention evaluations [11], we labeled reviews as high quality if they scored 14-18 on the OQAQ. Those falling slightly below this threshold (11-13) were labeled medium quality 11-13, and reviews scoring below 11 were regarded as low quality.

Data Extraction

From each included review, we extracted information concerning setting and methods (eg, country, context, study design, inclusion and exclusion criteria), participants (eg, total number of participants, missing participants, mean age, gender), outcome measures (method of assessing outcomes, duration), main findings (especially effectiveness summaries and analyses relating intervention content to effectiveness), intervention (eg, type of intervention, change targets in terms of cognitive, emotional, or physiological changes targeted and or assessed in process evaluations, mode of delivery, intervention content). Data extraction forms are available from the authors.

Analyses

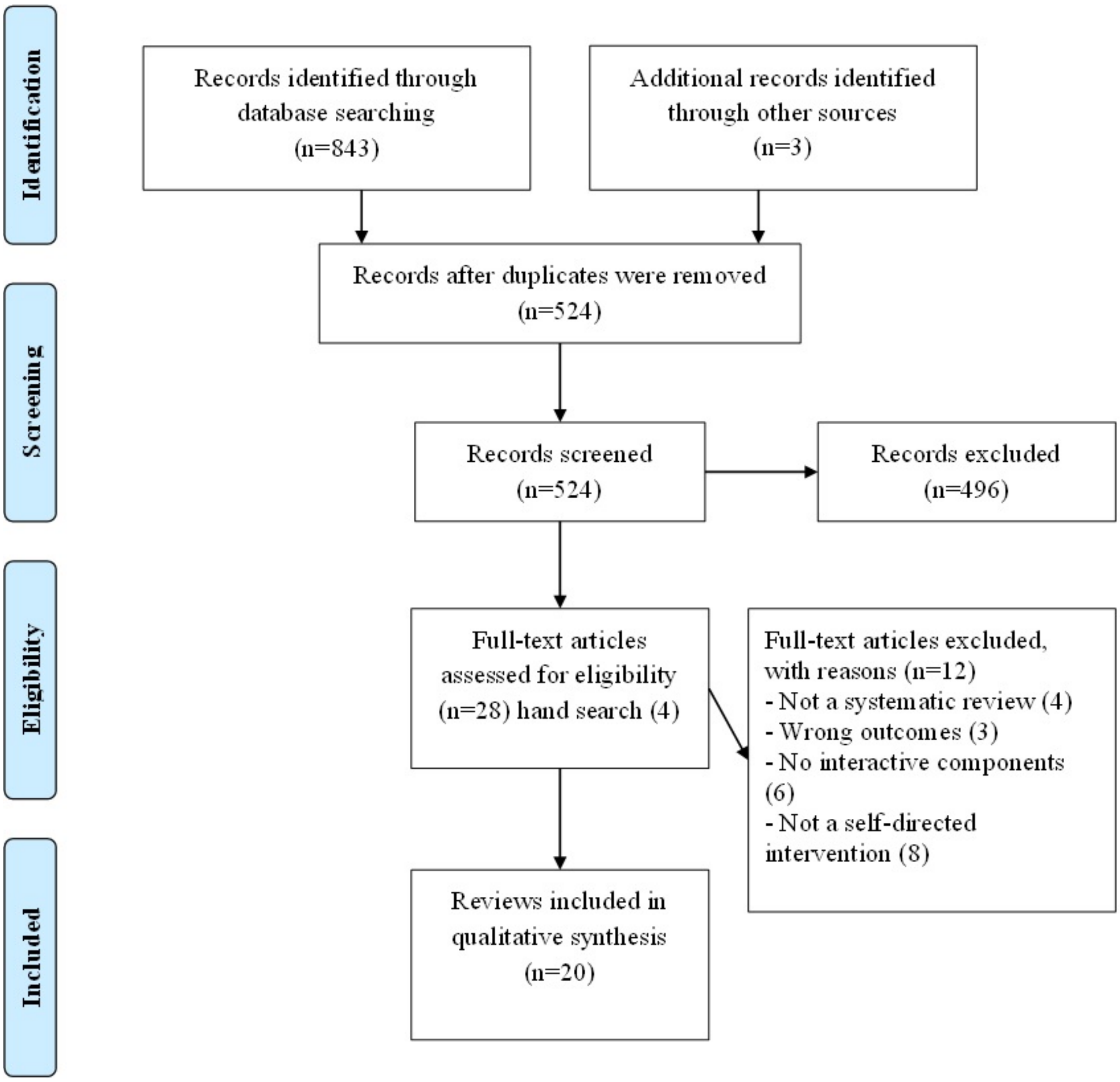
Each review was searched for descriptions of content of relevant self-directed weight loss interventions and for both statistical and narrative assessment of the relationship between intervention content and effectiveness. This information was extracted and is summarized for each review in Multimedia Appendix 2.

Results

Search Results

In accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, Figure 1 shows that our search strategy identified 524 articles after removal of duplicates. Title and abstract examination and reference-checking generated 32 potentially eligible articles of which 20 met our inclusion criteria. The number of participants included in studies within each review ranged from 298 to 12,417. Three of the selected reviews did not report total sample size.

Figure 1. Study selection; PRISMA flow diagram.



Review Characteristics

Included reviews were published between 2006 and 2012 and focused on weight loss in overweight or obese adults aged 20-79 years old. A summary description of each of the 20 reviews included is provided in Multimedia Appendix 2. The appendix describes review type, search period, inclusion criteria, OQAQ score, review aims, the number of primary intervention studies reviewed, the number of primary intervention studies meeting our inclusion criteria, the content of interventions meeting our inclusion criteria, and the overall results and conclusions.

Three of the selected reviews did not report total sample size across reviewed studies. Of the remaining, Enwald and Huotari [24] included the most participants (n=12,417) and Keller et al [25] the fewest (n=298). Most reviews focused on participants living in the community, although one review included home-based interventions for postpartum women [25].

All reviews, apart from one, summarized evaluation studies of weight loss or weight maintenance interventions. Three reviews excluded weight maintenance interventions [26-28]. Reviews included a variety of primary intervention evaluations, some relevant to our review and others not. For example, Kroeze et al reviewed the effectiveness of computer-tailored educational interventions evaluated in relation to a series of health-related outcomes, including smoking cessation, diabetes, and asthma management [29]. Consequently, only 2 of 31 primary evaluation studies included in this review were relevant to our research questions and so met our inclusion criteria. By contrast, all studies included in Weinstein met our inclusion criteria [30]. Thus for some reviews, we focus on conclusions based on a minority of the primary evaluations included in the review.

Only 4 reviews reported meta-analyses of weight-related outcomes [20,26,31,32]. The remaining 16 reviews reported narrative syntheses of primary studies (see Multimedia Appendix 2 for overview of reviews) [12,24,25,27,28-30,33-41].

Review Quality

The overall methodological quality of included reviews was relatively poor. The average OQAQ score was 12.8. Individual review scores are shown in Multimedia Appendix 2. Only 5 reviews applied study quality assessment criteria to inform their analyses/interpretations [20,26,31,37,40], and most reviews did not assess the methodological quality of primary studies or consider potential reporting biases. Of the 20 included reviews, Loveman et al [37], Tuah et al [40], and Wieland et al [20] had the highest quality scores (of 18), and 4 others were high quality, scoring 17 [26,31], 15 [32], and 14 [33].

Apart from Neve et al [32], who did not use quality assessment criteria, all high-quality reviews examined the methodological quality of primary studies. Four reviews used the Cochrane Collaboration Risk of Bias Tool [20,26,31,40]. Harris et al also used the Effective Public Health Practice Project quality assessment index [31], and 2 reviews developed their own methodological quality assessments [33,37]. While all 7 high-quality reviews included primary evaluations of interventions for overweight or obese adults, Cole-Lewis and Kershaw [33] and Harris et al [31] also included interventions with adolescents, so their conclusions do not refer exclusively to adult populations.

Weight Loss Effectiveness and Mode of Delivery

Across 20 reviews, we identified 99 primary evaluation studies that met our inclusion criteria. The interventions described in these studies employed a range of delivery formats including online programs, mobile phone applications, text messaging, email, electronic and print newsletters, telephone-based communication, print manuals, and booklets.

To explore findings, we grouped reviews according to the main delivery formats used by the interventions they considered. Nine reviews focused on Internet interventions. Three reviews evaluated interventions based on electronic devices such as mobile phones (referred to as “eHealth” interventions). Seven reviewed various multicomponent interventions, some of which were described as “home-based”, and one reviewed text-messaging interventions.

Internet-Based Interventions

In a narrative review judged to be of medium quality, Weinstein included 8 evaluative studies (5 assessing weight loss and 3 assessing weight maintenance) [30]. All 8 met our inclusion criteria. The review included data from 418 overweight or obese participants aged 30-62 years with intervention durations of 6-12 months. Four of the five weight loss studies supported Weinstein’s conclusion that Internet-based interventions could provide an alternative to traditional interventions achieving weight loss of 1.7 kg (SD 2.7) to 2.2 kg (SD 2.6). The exception compared two self-delivered approaches. This study found that participants using a manual-based program lost substantially more weight than those using a tailored online dietary intervention. Findings were equivocal for the 3 weight maintenance interventions, and Weinstein called for further research.

Weinstein concluded that the content of Internet-based interventions is crucial to effectiveness and highlighted the potential importance of use of food records, sending weekly emails, and using telephone reminders. Weinstein called for further randomized trials on the use of Internet-based interventions for weight loss and weight maintenance [30].

Harvey-Berino et al randomized 250 participants to an Internet support group, minimal in-person support, or frequent in-person support group following a 6-month weight loss intervention involving interactive television. After 12 months, no differences were found. Interestingly, the Internet-based group experienced greater peer contact and were more likely to complete self-monitoring diaries but nonetheless had the highest attrition rate, suggesting that, for some participants, the intervention had diminishing appeal over time [42].

In a medium quality review, Kroeze et al report effectiveness of computer-tailored education on physical activity and dietary behaviors [29]. Only 2 of the 31 studies met our inclusion criteria because, although many assessed weight-related outcomes, most did not report weight lost. Results were inconclusive in relation to weight loss effectiveness.

In a narrative review judged to be low quality, Saperstein et al included 6 studies of online social support interventions that included feedback from a therapist, e-bulletin boards for peer support, and email communication with a counselor [28]. All 6 studies met our inclusion criteria. These interventions were effective with interventions achieving a weight loss range of 2.6-8.3kg, but only when specific change strategies were used. Information provision alone without feedback had no effect on weight outcomes. Saperstein et al concluded that “personalization through ongoing tailored information and feedback, either via email from a human counsellor or a computer-based program, was a critical component” (p. 4).

In another narrative review judged to be of medium quality, Turk et al included 40 studies that focused primarily on weight maintenance [41]. Only 8 of these studies met our inclusion criteria. Findings from 2 of these suggest that behavioral interventions with online chat sessions delivered via the Internet were as effective as an in-person behavioral therapy intervention [42,43]. Contrary to these findings, however, Harvey-Berino et al reported that an Internet chat group maintained significantly less weight than a minimal in-person and frequent in-person group (-5.7 kg [SD 5.9] vs -10.4 kg [SD 9.3] vs -10.4 kg [SD 6.3], respectively) [44]. Moreover, in Wing et al [45], an Internet chat room group was less successful in maintaining weight than in-person behavioral treatment (4.7 kg [SD 8.6] vs 2.5 kg [SD 6.7], respectively) [45]. These studies describe different interventions evaluated using different methods. The interventions evaluated in these studies vary greatly methodologically, which renders data synthesis impossible. Thus, findings in this review are equivocal, and the relative effectiveness of Internet versus face-to-face groups warrants further investigation.

Neve et al reported a high-quality meta-analytic review [32]. All 18 studies met our inclusion criteria. Random effects meta-analysis of 3 studies demonstrated a significant difference between an enhanced Web-based intervention (involving self-monitoring activities and individual email feedback) and an education-only Web-based intervention with less weight regained in the enhanced Web-based intervention group post intervention (weighted mean difference 2.24; 95% CI 1.27-3.21; I^2=20.9%) [46-48]. Two weight maintenance studies combined in a meta-analysis also demonstrated less weight regain in participants involved in a Web-based intervention compared to a minimal intervention or usual care control group (weighted mean difference -0.30; CI -0.34 to -0.26; I^2=0%) [49,50]. Although these meta-analyses supported the efficacy of Internet interventions, only 3 of the studies reported to achieve clinical weight loss of 5%. Neve et al were unable to draw generalizable conclusions due to the small numbers of comparable interventions [32].

In a narrative review judged to be of medium quality, Manzoni et al [38] updated Neve et al's review [32] of 18 Internet-based studies, including 8 additional studies. All studies met our inclusion criteria and focused primarily on teenage women. Interventions lasted from 6 weeks to 2 years. Manzoni et al confirmed previous findings suggesting that Internet interventions including feedback are more effective than those providing information alone [38]. For example, Bennett et al randomized 101 participants to an Internet-based intervention including counseling sessions, behavior change goals, and self-monitoring compared to "usual care". After 3 months, the intervention group lost 2.3 kg compared to a gain of 0.28 kg in the usual care group [51]. Overall, however, Manzoni et al judged available evidence to be inconclusive because of heterogeneity in duration and intensity of interventions, and variation in the type of feedback and social support tools used [38].

Manzoni et al also attempted to assess the cost-effectiveness of Internet-based interventions [38]. Two studies suggested that Internet-based interventions could be effective and less expensive than alternative interventions [52,53]. For example, Booth et al estimated that, over 12 weeks, Internet-based interventions could save up to US $155 compared to in-person interventions [52]. In addition, the authors reported further savings in travel time and travel costs after the 12-week period. However, only 2 studies provided cost-effectiveness data. Thus, further studies are needed to clarify how much less expensive Internet-based interventions could be when compared to standard weight loss interventions.

In a high-quality review, Reed et al identified 11 RCTs of Internet weight loss interventions, including email counseling and handheld, self-monitoring computer devices [26]. Seven of these studies, focusing on teenage women, met our inclusion criteria. Random effects meta-analyses of 6 of these studies found that adding a computer-based intervention to standard weight loss treatment significantly increased the amount of weight lost between 2 and 12 months (weighted mean difference −1.48 kg, 95% CI −2.52 to -0.43; I^2=0%; P=.01) [54-58]. However, 4 of these studies were not primarily self-delivered interventions and so did not meet our inclusion criteria [54,55,57]. A meta-analysis of 5 exclusively self-directed computer-based interventions found that computer-based groups lost less weight than the standard treatment groups (using paper-based materials) (weighted mean difference 1.47 kg, 95% CI 0.13-2.81; I^2=0%; P<.001) [59-62]. The authors concluded that Internet-based interventions were effective only when used in conjunction with a standard weight loss intervention, but the amount of weight lost (<1.5 kg) was too small to be clinically relevant for overweight and obese populations.

In a low-quality narrative review, Arem et al reviewed 9 RCTs, 8 of which met our inclusion criteria. These 8 studies reported modest weight loss ranging from 0.8-4.9 kg across studies [12]. For example, Hunter et al reported a 1.3 kg (SD 4.1) weight loss in 446 military participants who took part in an Internet-based intervention compared to a weight gain of 0.6 kg (SD 3.4) for those in a "usual care" group [63]. Rothert et al randomized 2862 participants to an online personalized information group and an online information-only group. After 6 months, the online personalized group lost significantly more weight (2.8 kg [SD 0.3] vs 1.1 kg [SD 0.4], respectively) [64]. Nonetheless, Arem et al [12] judged the data to be inconclusive due to the high attrition rates and variability/incompatibility of intervention methods. The most promising interventions identified in this review were in-person treatments followed by online weight maintenance, and only this combination achieved 5% weight loss. It is unclear, therefore, how effective Internet weight loss interventions are without personal contact/support.

Wieland et al reported a high-quality review examining the effectiveness of interactive computer-based interventions on weight loss and weight maintenance in obese or overweight adults [20]. All 18 studies reviewed met our inclusion criteria. These studies involved 4140 participants from the community, and 14 included weight loss as a primary outcome. For example, at 3 months, a meta-analysis of five weight loss trials found that computer-based intervention participants lost more weight than those in the minimal control group (eg, printed material or no treatment conditions; weighted mean difference -2.5 kg; 95% CI -3.4 to -1.6; I^2=44%) [51,56,65-67]. Four studies focused on weight maintenance (as opposed to initial weight loss) [43,45,49,50]. Participants using the computer-based interventions regained less weight than those in the minimal treatment or no treatment control at 6 months (mean difference -0.7 kg; 95% CI -1.2 to -0.2; two trials) [45,50] and at 12 months (mean difference -0.8 kg; 95% CI -1.4 to -0.2; three trials) [45,49,50]. One trial compared a computer-based intervention to an intensive in-person intervention (involving contact every 2 weeks over 12 months). Participants in the computer-based intervention regained more weight during the first 6 months (weighted mean difference 2.2 kg; 95% CI 0.3-4.1), and at 12 months lost less weight than the in-person group (weighted mean difference 4.7 kg; 95% CI 1.7-7.7) [43]. Overall, the authors of this high-quality meta-analytic review concluded that, compared to no intervention or minimal interventions, computer-based interventions are effective in prompting weight loss and in supporting maintenance of weight loss. However, computer-based interventions result in less weight loss and greater weight regain than in-person interventions. So, for example, while computer-based interventions may result in approximately 2.5 kg loss over 3 months, in-person interventions can achieve up to 10% of weight loss at up to 26 months [68].

eHealth Interventions

Reviews used the term "eHealth" to refer to interventions delivered using electronic devices including smartphones and Internet-based computer interventions. Thus, the 3 reviews considered in this delivery category include primary evaluation studies that could also have been included in reviews of Internet intervention studies (as above).

In a narrative review judged to be of medium quality, Norman et al summarized 49 eHealth and Internet studies [39]. All studies targeted both dietary and physical activity behavior change, and interventions lasted 4-12 months. These studies were of variable quality, and only 12 met our criteria with 33 failing to include a weight-related outcome. Norman et al drew few conclusions but recommended that future studies focus on underlying mechanisms and change techniques that promote dietary and physical activity behavior change.

In a narrative review judged to be of medium quality, Enwald and Huotari evaluated electronic interventions for the prevention of obesity and its associated health problems [24]. Of the 23 included studies, 21 were RCTs and 2 employed quasi-experimental designs. However, only 5 studies met our inclusion criteria with 17 failing to include a weight-related outcome. Interventions included emails, use of websites, electronic feedback, CD-ROM, and newsletters and lasted between 1 and 12 months. Results showed that dietary interventions had a greater influence on weight than physical activity programs. Interestingly, tailoring was reported to be more effective when applied in the context of dietary interventions and less effective in physical activity programs.

By contrast, Harris et al conducted a high-quality meta-analytic review of 43 studies, of which 22 met our inclusion criteria; 21 did not include weight loss outcomes [31]. Across 40 adult studies participant ages ranged from 40-49 years. 27 of these studies used the Food Frequency Questionnaire, and others used a variety of outcomes. Interventions lasted between 1 and 6 months, and participants interacted with the intervention either daily or weekly for 10-45 minutes. Based on dietary behaviors outcomes, in particular intake of fruit, vegetable, fat, and fiber, this review found no evidence that eHealth was more effective or cost-effective than in-person interventions. Four self-delivered interventions reported mean weight [48,69-71], and three reported mean change in weight [49,50,58]. Random effects meta-analyses of both groups, that is, (1) the former four (weighted mean difference 0.6 kg; 95% CI -3.5 kg to 4.6 kg; *P*=.78) and (2) the latter three, found no evidence of intervention effect (weighted mean difference –0.07 kg; 95% CI –1.8 kg to 1.6 kg; *P*=.94). However, large heterogeneity of effect sizes casts some doubt on the applicability of these average results across intervention evaluations. In conclusion, however, this high-quality meta-analytic review found no evidence of effectiveness of self-delivered eHealth weight loss interventions in comparison with other approaches.

Home-Based Print and Multicomponent Interventions

Three reviews summarized intervention evaluation studies, many of which were referred to as “home-based” [25,27,35]. These, together with interventions reviewed by 4 other reviews, typically included mailed instructions or advice on dietary and physical activity (eg, brochures, leaflets, health professional advice), self-monitored physical activity using electronic devices (eg, pedometers, accelerometers), promotion of diaries, and provision of various written materials—or a combination of the above [34,36,37,40].

In a narrative review judged to be low quality, Hemmingsson et al included 7 studies of physical activity, 2 of which met our inclusion criteria [35]. For example, Perri et al compared a “home-based”, individual walking intervention with an organized group-based walking intervention. At 12 months, participants in the individual intervention reported 20.8 minutes more walking per week than those in the group intervention. Those in the home-based group also lost more weight after 15 months (11.65 kg [SD 8.99] vs 7.01 kg [SD 8.23]) [72]. However, this was based only on a small sample of 49 obese women enrolled in a behavioral modification program. Further investigation of the potential of individual walking interventions is warranted.

In a narrative review of weight management interventions for postpartum women, judged to be low quality, Keller et al included 6 studies of which 3 met our inclusion criteria [25]. All 6 demonstrated significant changes in body composition with a reported weight loss range of 1.6-7.8 kg in 3 studies. For example, Leermakers et al found that the behavioral weight loss intervention group involving telephone contact and 16 written lessons on exercise, nutrition, and behavior change strategies lost more weight than the no treatment control group involving healthy eating and exercise informational brochures, after 6 months (7.8 kg vs 4.9 kg, respectively) [73].

In another narrative review judged to be low quality, Lemmens et al included 9 studies of interventions with adults of which 3 met our inclusion criteria [27]. The 9 interventions included home-based exercises, written materials, emails, and face-to-face sessions. Only one of these found a small but statistically significant weight loss difference of 1.6 kg.

In a high-quality narrative review, Loveman et al included 12 studies of multicomponent interventions that involved home-based weight loss schemes [37]. Of these, 10 met our inclusion criteria. Many of these studies reported small, average weight loss. However, variability in intervention duration, intensity, addition of subsequent weight maintenance intervention components, and length of follow-up prevented drawing of meaningful conclusions regarding common elements associated with effectiveness.

In a low-quality narrative review focused on weight gain prevention interventions, Lombard et al included 9 studies [36]. In general, low intensity multicomponent interventions combining physical activity, diet, and behavior change content were found to be effective for preventing weight gain. Weight loss range was 1-1.9kg for 7 studies matching our inclusion criteria. However, only 5 demonstrated significant findings. For example, in another study, Lombard et al examined an intervention consisting of four group-based behavior change sessions followed by text messages and monthly mail contact over a 1-year period. They found a difference of -1.01 kg (*P*=.03) of weight loss between the intervention and a control group (involving group-based education sessions) [74]. Overall, only a few studies assessed the effectiveness of interventions designed to prevent weight gain, and like other multicomponent reviews, intervention content varied across trials making it difficult to compare effect sizes and to generate robust conclusions.

Gordon et al reported a low-quality narrative review focusing on pharmaceutical and in-person weight loss treatment [34]. Two of the 10 included studies were primarily self-delivered and matched our inclusion criteria. For example, Ahrens et al compared an intervention incorporating personalized information sheets and tailored exercise advice to a reduced calorie diet group among 95 participants. After 6 months, no significant difference in weight loss was reported between groups [75]. Again variability across interventions made it difficult to draw conclusions about intervention components associated with effectiveness.

In a narrative review judged to be high quality, Tuah et al identified studies that applied the transtheoretical model (TTM) to weight loss, but only 2 of the 5 studies in this review were primarily self-delivered [40]. While these 2 studies reported a small change in weight, this was not sustained over 24 months. The authors concluded that “trials that used stages of change as an assessment and intervention framework, rather than just as a tool to assign and assess stage of change, reported minimal weight loss” (p. 18).

Text Message Interventions

In a high-quality narrative review, Cole-Lewis and Kershaw summarized 12 studies of SMS text messaging (short message service, SMS) interventions promoting a range of health behaviors including smoking cessation, diabetes, and asthma management [33]. Only 2 of these studies met our inclusion criteria. Both reported effective text messaging interventions with a weight loss range of 2.9-4.5kg. For example, Haapala et al randomized 126 overweight adults aged 25-44 years to a text message or a no-contact control group. After 12 months, the intervention group lost more weight than the control group (4.5 kg/m^2 vs 1.1 kg/m^2, P=.006, respectively) [76]. Weight loss occurred mostly in the first 3 months when usage of the text message program was high, so the longer-term effects of text messaging were unclear.

Change Mechanisms and Theoretical Frameworks

None of the 20 reviews drew conclusions regarding the usefulness of particular theories or mechanisms of change. However, some reviews did highlight theories underpinning intervention design.

Enwald and Huotari reported that the most commonly mentioned theory in the evaluation studies in their review was the transtheoretical model (TTM), which guided 14 of 23 studies [24]. Other theories used included the Elaboration Likelihood Model, the Precaution Adoption Model, the Theory of Reasoned Action, the Theory of Planned Behavior, Goal Setting Theory, and the Health Promotion Model. Enwald and Huotari did not relate the theoretical foundation of interventions to effectiveness.

Tuah et al identified two interventions applying TTM to weight loss, both of which resulted in small losses in weight that were not sustained over 24 months [40]. The authors reported that TTM-based interventions using feedback, self-monitoring, anthropometric measurements, and counseling resulted in significant effects on weight loss.

Harris et al considered use of theory and change mechanisms in relation to changes in fruit, vegetable, fiber, and fat intake [31]. Of the 13 effective interventions, they found only one study that employed theory to identify change mechanisms. Anderson et al reported that self-efficacy and outcome expectancies in relation to physical activity mediated greater consumption of fruit, fiber, vegetables, and fat [77]. However, no meditational analyses were conducted.

Change Techniques and Delivery Formats

Only Wieland et al provided meta-analyses linking specific intervention components with effectiveness [20]. At 3 months, meta-analysis of 3 trials demonstrated that participants receiving Internet-based interventions supplemented with individualized feedback experienced greater weight loss than participants in an Internet-based intervention without individualized feedback (weighted mean difference -2.1 kg; 95% CI -2.9 to -1.4; P<.001) [46,48,78]. A similar effect was also found for participants using email counseling in 3 trials (weighted mean difference -2.3 kg; 95% CI -3.1 to -1.5; P<.001) [46,48,78] and automated feedback in one trial (weighted mean difference -1.8 kg; 95% CI -3.2 to -0.5; P=.009) [48] when compared to an Internet-intervention delivered alone.

At 3 months, non-directive email counseling did not induce significant weight loss in one trial (weighted mean difference –0.3 kg; 95% CI –2.2 to 1.7; P=.80) [78]. Similarly, no effect was found at 4 months for a group chat intervention delivered in conjunction with online self-monitoring (weighted mean difference 1.5 kg; 95% CI -0.7 to 3.7; P=.18). However, this was based on only one trial with no follow-up assessment beyond 4 months [79]. The authors concluded computer-based feedback delivered in conjunction with an Internet-based intervention enhances weight loss.

Brief descriptions of the intervention content identified in narrative reviews are included in Multimedia Appendix 2. The most commonly mentioned mechanism-based change “techniques” [17] across reviews were self-monitoring, feedback, and goal setting. Interventions including these change techniques were generally more effective than information only interventions. Reminders were used in a number of effective interventions as were self-efficacy enhancement techniques [18] and provision of counseling opportunities.

Manzoni et al noted that most effective interventions promoting weight loss and maintenance incorporated tailored feedback via email, e-counseling, food diaries, and self-monitoring of physical activity, diet, and weight [38]. However, in the absence of meditational analyses, it is unclear which combination of techniques and delivery formats enhanced effectiveness. Lombard et al observed that “self-monitoring of weight was a component of three [effective] interventions. Four interventions used self-monitoring of diet or physical activity, but the form or reason for monitoring was not always clear” (p. 2243) [36].

Neve et al explored intervention components within individual studies and concluded that social support, peer support contact, and online bulletin boards increased website usage but no meditational analyses were presented [32]. Similarly, Weinstein noted that effective interventions included “social” components such as e-counseling from a therapist and an online bulletin, but whether these components are directly linked to effectiveness requires further investigation [30].

Behavior change techniques were delivered by means of a variety of “delivery formats” including Internet sites, emails, text messaging, CDs, telephone calls, pedometers, paper questionnaires and diaries, manuals, pamphlets, booklets, brochures, and workbooks. However, reviews did not allow firm conclusions to be drawn regarding these delivery formats. Internet programs appear to be effective, especially in comparison with no intervention or minimal-contact interventions and have the capacity to enhance the effectiveness of in-person programs. Personal tailoring of programs may also enhance the effectiveness of self-directed weight loss interventions.

Discussion

Principal Findings

To our knowledge, this is the first systematic review of reviews to examine the effectiveness of self-directed weight loss interventions. Twenty reviews including 99 primary evaluations met our inclusion criteria. Only 7 of 20 reviews were high quality according to criteria specified by OQAQ but given the paucity of available evidence, we summarized evidence from all 20 reviews. The reviews identify a variety of potentially effective, self-directed weight loss interventions delivered by means of the Internet, mobile electronic devices, print media, and combinations of these delivery formats.

Three reviews focusing on Internet-based interventions and one focusing on eHealth interventions conducted meta-analyses to determine which intervention type, duration, and intensity were the most effective. Of these 4 meta-analytic reviews, the strongest evidence comes from Wieland et al, where all studies met our inclusion criteria [20]. For example, a meta-analysis of 5 trials demonstrated that self-delivered interactive computer-based programs were more effective than minimal interventions (eg, printed newsletters) or no treatment, for short-term weight loss and weight maintenance. This finding corresponds to that reported by Neve et al [32]. However, most trials included in these reviews did not examine weight outcomes beyond 1-year follow-up, so the impact of computer or Internet-based interventions on long-term weight loss is unclear.

Our first research question concerned the effectiveness of self-directed weight loss interventions. Reed et al [26] and Wieland et al [20] suggest that computer or Internet-based interventions are less effective than in-person treatment, but further trials are needed to clarify whether the greater weight loss observed following in-person treatments is replicable, clinically significant, and cost effective. Reed et al concluded that computer-based interventions delivered in conjunction with standard treatment enhance weight loss compared to standard treatment delivered alone. However, this meta-analysis included just 6 trials and the magnitude of weight lost advantage was small (<1.5 kg).

Overall, weight loss (kg) across all relevant studies reported in 7 reviews ranges from 0.8-7.8kg. Caution is advised in the interpretation of these figures as wide variations were found in intervention content and delivery. For example, most reviews did not report intervention intensity and frequency and of those that did, this varied from 1 week to 1 year. Five Internet-based studies within 2 reviews achieved a percentage weight loss of 5% [12,32], which has been used as a benchmark associated to health benefits [80,81]. However, most reviews did not report whether interventions achieve 5% weight loss. Therefore, it is unclear how many interventions achieved clinically significant weight loss.

Enhanced Web-based intervention involving self-monitoring activities and email feedback appear to be more effective than information-provision alone, but this conclusion was supported by only one meta-analysis including only 3 primary studies including 375 participants [32]. Currently no meta-analyses have been conducted examining text message, home-based print, and multicomponent delivery formats. Two primary studies reviewed by Cole-Lewis and Kershaw present promising findings in relation to short-term weight loss following text messaging interventions [33]. Narrative reviews of home-based print and multicomponent delivery formats are inconclusive, in part because of the heterogeneity of content found across these interventions.

Our second research aim was to investigate whether effectiveness is enhanced by inclusion of particular behavior change techniques. We found that underlying components within self-directed interventions that contribute to weight loss success were largely unexplored at both study and review level. No meta-analyses were available assessing associations between included techniques and weight loss. Reviewing single trials, Wieland et al observed that Internet-based interventions including individualized feedback or email counseling had been found to be more effective than Internet-based interventions that did not employ these techniques [20]. Conversely, interventions including non-directive email, group chat, and online self-monitoring had been found to be less effective than Internet-based interventions that did not employ these techniques.

Narrative reviews described the content of effective interventions and found that these tended to employ self-monitoring, feedback, and goal setting. However, these reviews did not examine whether such techniques were associated with enhanced weight loss or weight maintenance. No review presented evidence on dose-response data for included change techniques, so it remains unclear whether mere inclusion or frequency of technique use is important to efficacy.

Our third research aim was to investigate whether effectiveness is enhanced by using particular delivery formats. We found that definitive conclusions could not be drawn regarding the most effective delivery format for self-directed weight loss interventions. This may depend on target audience. A greater number of primary studies of Internet-based interventions are available, and these are found to be more effective than minimal interventions (such as provision of leaflets). Trials of other delivery formats, such as eHealth interventions and text messaging, suggest that such interventions can be effective. The advantages of all such interventions include personal tailoring of information, 24-hour availability, anonymity, online social support, and affordability. These characteristics imply that, when effective, such interventions are likely to be cost-effective. Unfortunately, many available evaluation studies are pilot or efficacy trials, rather than definitive trials, and few cost-effectiveness studies have been undertaken. Further evaluation studies using large samples with long-term weight loss follow up and cost-effectiveness analyses are needed.

Strengths and Limitations

Our review identified a range of reviews including primary evaluation studies of self-directed interventions designed to reduce weight. Every effort was made to reduce bias in the search, selection of reviews, data extraction, and data analysis. This review provides an overview of what is currently known in this rapidly expanding research area.

Nonetheless, several challenges affecting our selection and interpretation of available evidence must be acknowledged. We relied on descriptions of interventions provided by reviewers, and these varied considerably in form and detail. Reviews also varied in their methods and in the quality of the review methodology employed with only 7 of 20 scoring highly on the OQAQ. In addition, the literature may well contain more primary evaluation studies that would meet our inclusion criteria than the 99 included in these 20 reviews. Furthermore, we found no reviews that used meta-regression [16] to examine relationships between intervention content and weight loss effectiveness, so suggestions rather than conclusions emerged in relation to our second and third research questions concerning the association between inclusion of particular behavior change techniques and use of particular modes of delivery and weight loss.

What Further Research is Needed?

Comprehensive Review of Primary Evaluations

A comprehensive review of primary evaluations of self-directed weight loss intervention evaluation studies is needed. Such a review would capture studies beyond the 99 primary evaluations included in our 20 reviews. Such a review should compare interventions using similar delivery formats, taking account of the potentially varying content of comparison groups and relate techniques and materials to effectiveness.

Such a review should use a quality assessment tool to assess review methods. Some reviews identified here used the Cochrane Collaboration Risk of Bias Tool. This comprises questions divided into seven areas: generation of the allocation sequence, concealment of the allocation sequence, blinding, attrition and exclusions, other generic sources of bias, biases specific to the trial design (crossover or cluster randomized trials), and biases. Only one review used the Effective Public Health Practice Project Tool, designed for use in public health and including questions concerning eight specific areas: selection bias, study design, confounders, blinding, data collection methods, withdrawals and dropouts, intervention integrity, and analyses. Both tools are useful. The latter may be more appropriate when reviewing large scale population intervention evaluation studies.

Further Meta-Analyses Focusing on Intervention Components

As part of a comprehensive review of primary evaluation studies of self-directed weight loss interventions, meta-analyses focusing on high-quality evaluation studies could be used to identify common intervention components in studies segmented by delivery format. This would generate quantitative answers to questions such as “What content works best for website-based weight loss interventions?” and “What content works best for weight loss mobile phone applications?” [82]. This would extend the work of the 4 meta-analyses identified in our review [20,26,31,32] and provide clear answers to the second and third research questions we addressed. In addition, meta-regression, controlling for co-occurrence of change techniques across interventions, could clarify whether theory-based combinations of techniques enhance weight loss effectiveness [16]. Finally such a review should consider the varying content of comparison conditions (such as usual care or alternative interventions), as such control content has demonstrable effects on the observed efficacy of interventions [83,84].

Further High-Quality Primary Evaluations Comparing Different Modes of Delivery

Further high-quality primary evaluations that compare different modes of delivery for the same (or very similar) interventions within particular populations are needed. Such studies should be reported in accordance with CONSORT guidelines [85] and include lists and specification of behavior change techniques included in the intervention content design. Results could recommend whether particular approaches such as Internet site, text messages, or mixed methods home delivery are most likely to be effective. Such studies should clearly specify the nature of comparison groups, specifying what constitutes control conditions.

Further High-Quality Primary Evaluations Using Objective Measures of Weight Loss at Longer-Term Follow-Up

Further high-quality primary evaluations that use objective measures of weight loss should be used at longer-term follow-up to assess maintenance of weight loss. These should follow the recommendations above and include multiple weight assessments lasting over 1 or, ideally 2 or more years. Such evaluations, conducted to scale, could provide population effectiveness data rather than the efficacy data on initiation of weight loss provided by most current intervention evaluation studies.

Conclusions

A systematic search identified 20 reviews including 99 primary evaluations of self-directed interventions designed to reduce weight. The evidence reviewed suggests that self-directed interventions can independently promote weight loss and can augment interventions involving personal contact. Some reviews identified techniques and delivery formats used in effective interventions, such as self-monitoring, feedback, self-efficacy enhancement, and social and peer support. However, it was not possible to infer which techniques or delivery modes are most strongly associated with increased weight loss for whom and in what contexts. Further primary evaluations of self-delivered weight loss interventions that clearly specify the behavior change techniques and materials employed are needed, especially with long-term follow-up. Further meta-analytic reviews focusing on weight loss intervention content and efficacy within delivery mode could provide better guidance for intervention designers and commissioners.

Acknowledgments

The listed order of the authors represents extent of contribution. The authors would like to thank Morwenna Rogers who helped develop the search strategy and Aleksandra Borek who helped check its application. The work was partially funded by the UK National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care of the South West Peninsula (PenCLAHRC), but the views expressed in this paper are those of the authors and not necessarily those of NIHR or the UK Department of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Selection of the final 20 reviews applying inclusion criteria.

[PDF File (Adobe PDF File), 58KB - jmir_v16i2e58_app1.pdf]

Multimedia Appendix 2

Review characteristics.

[PDF File (Adobe PDF File), 64KB - jmir_v16i2e58_app2.pdf]

References

1. Allender S, Peto V, Scarborough P, Boxer A, Rayner M. Coronary Heart Disease Statistics. London: British Heart Foundation; 2006.
2. Prospective Studies Collaboration. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. The Lancet 2009 Mar;373(9669):1083-1096. [doi: 10.1016/S0140-6736(09)60318-4]
3. 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]
4. Christensen R, Kristensen PK, Bartels EM, Bliddal H, Astrup A. Efficacy and safety of the weight-loss drug rimonabant: a meta-analysis of randomised trials. Lancet 2007 Nov 17;370(9600):1706-1713. [doi: 10.1016/S0140-6736(07)61721-8] [Medline: 18022033]
5. Hill JO, Peters JC. Environmental contributions to the obesity epidemic. Science 1998 May 29;280(5368):1371-1374. [Medline: 9603719]
6. Jeffery RW, Wing RR, Sherwood NE, Tate DF. Physical activity and weight loss: does prescribing higher physical activity goals improve outcome? Am J Clin Nutr 2003 Oct;78(4):684-689 [FREE Full text] [Medline: 14522725]
7. Lench HC, Flores SA, Bench SW. Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: a meta-analysis of experimental emotion elicitations. Psychol Bull 2011 Sep;137(5):834-855. [doi: 10.1037/a0024244] [Medline: 21766999]
8. Kessler D. The End of Overeating: Taking Control of the Insatiable American Appetite. New York: Rodale; 2009.
9. National Institute for Health and Clinical Excellence. Obesity: The prevention, identification, assessment and management of overweight and obesity in adults and children. In: NICE clinical guidelines, No. 43. London: Centre for Public Health Excellence at NICE; National Collaborating Centre for Primary Care; Dec 2006.
10. Muñoz RF, Mendelson T. Toward evidence-based interventions for diverse populations: The San Francisco General Hospital prevention and treatment manuals. J Consult Clin Psychol 2005 Oct;73(5):790-799. [doi: 10.1037/0022-006X.73.5.790] [Medline: 16287379]
11. 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]
12. Arem H, Irwin M. A review of web-based weight loss interventions in adults. Obes Rev 2011 May;12(5):e236-e243 [FREE Full text] [doi: 10.1111/j.1467-789X.2010.00787.x] [Medline: 20804523]
13. 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]
14. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 2;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]

15. Muñoz RF, Lenert LL, Delucchi K, Stoddard J, Perez JE, Penilla C, et al. Toward evidence-based Internet interventions: A Spanish/English Web site for international smoking cessation trials. Nicotine Tob Res 2006 Feb;8(1):77-87. [doi: 10.1080/14622200500431940] [Medline: 16497602]
16. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
17. 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]
18. Abraham C. Mapping change mechanisms and behavior change techniques: A systematic approach to promoting behavior change through text. In: Abraham C, Kools M, editors. Writing Health Communication: An Evidence-Based Guide. London: SAGE Publications Ltd; 2012:99-115.
19. Davidson KW, Goldstein M, Kaplan RM, Kaufmann PG, Knatterud GL, Orleans CT, et al. Evidence-based behavioral medicine: what is it and how do we achieve it? Ann Behav Med 2003 Dec;26(3):161-171. [Medline: 14644692]
20. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012;8:CD007675. [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
21. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
22. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
23. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991;44(11):1271-1278. [Medline: 1834807]
24. Enwald HP, Huotari ML. Preventing the obesity epidemic by second generation tailored health communication: an interdisciplinary review. J Med Internet Res 2010;12(2):e24 [FREE Full text] [doi: 10.2196/jmir.1409] [Medline: 20584698]
25. Keller C, Records K, Ainsworth B, Permana P, Coonrod DV. Interventions for weight management in postpartum women. J Obstet Gynecol Neonatal Nurs 2008;37(1):71-79. [doi: 10.1111/j.1552-6909.2007.00202.x] [Medline: 18226159]
26. Reed VA, Schifferdecker KE, Rezaee ME, O'Connor S, Larson RJ. The effect of computers for weight loss: a systematic review and meta-analysis of randomized trials. J Gen Intern Med 2012 Jan;27(1):99-108 [FREE Full text] [doi: 10.1007/s11606-011-1803-9] [Medline: 21805218]
27. Lemmens VE, Oenema A, Klepp KI, Henriksen HB, Brug J. A systematic review of the evidence regarding efficacy of obesity prevention interventions among adults. Obes Rev 2008 Sep;9(5):446-455. [doi: 10.1111/j.1467-789X.2008.00468.x] [Medline: 18298429]
28. 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]
29. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Ann Behav Med 2006 Jun;31(3):205-223. [doi: 10.1207/s15324796abm3103_2] [Medline: 16700634]
30. Weinstein PK. A review of weight loss programs delivered via the Internet. J Cardiovasc Nurs 2006;21(4):251-8; quiz 259. [Medline: 16823276]
31. Harris J, Felix L, Miners A, Murray E, Michie S, Ferguson E, et al. Adaptive e-learning to improve dietary behaviour: a systematic review and cost-effectiveness analysis. Health Technol Assess 2011 Oct;15(37):1-160 [FREE Full text] [doi: 10.3310/hta15370] [Medline: 22030014]
32. 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]
33. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010 Apr;32(1):56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
34. Gordon J, Watson M, Avenell A. Lightening the load? A systematic review of community pharmacy-based weight management interventions. Obes Rev 2011 Nov;12(11):897-911. [doi: 10.1111/j.1467-789X.2011.00913.x] [Medline: 21883869]
35. Hemmingsson E, Page A, Fox K, Rossner S. Influencing adherence to physical activity behavior change in obese adults. Scandinavian Journal of Nutrition 2001;45:114-119.
36. Lombard CB, Deeks AA, Teede HJ. A systematic review of interventions aimed at the prevention of weight gain in adults. Public Health Nutr 2009 Nov;12(11):2236-2246. [doi: 10.1017/S1368980009990577] [Medline: 19650959]
37. Loveman E, Frampton GK, Shepherd J, Picot J, Cooper K, Bryant J, et al. The clinical effectiveness and cost-effectiveness of long-term weight management schemes for adults: a systematic review. Health Technol Assess 2011 Jan;15(2):1-182 [FREE Full text] [doi: 10.3310/hta15020] [Medline: 21247515]

38. Manzoni GM, Pagnini F, Corti S, Molinari E, Castelnuovo G. Internet-based behavioral interventions for obesity: an updated systematic review. Clin Pract Epidemiol Ment Health 2011;7:19-28 [FREE Full text] [doi: 10.2174/1745017901107010019] [Medline: 21552423]
39. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
40. Tuah NA, Amiel C, Qureshi S, Car J, Kaur B, Majeed A. Transtheoretical model for dietary and physical exercise modification in weight loss management for overweight and obese adults. Cochrane Database Syst Rev 2011(10):CD008066. [doi: 10.1002/14651858.CD008066.pub2] [Medline: 21975777]
41. Turk MW, Yang K, Hravnak M, Sereika SM, Ewing LJ, Burke LE. Randomized clinical trials of weight loss maintenance: a review. J Cardiovasc Nurs 2009;24(1):58-80 [FREE Full text] [doi: 10.1097/01.JCN.0000317471.58048.32] [Medline: 19114803]
42. 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]
43. Harvey-Berino J, Pintauro SJ, Gold EC. The feasibility of using Internet support for the maintenance of weight loss. Behav Modif 2002 Jan;26(1):103-116. [Medline: 11799651]
44. Harvey-Berino J, Pintauro S, Buzzell P, DiGiulio M, Casey Gold B, Moldovan C, et al. Does using the Internet facilitate the maintenance of weight loss? Int J Obes Relat Metab Disord 2002 Sep;26(9):1254-1260 [FREE Full text] [doi: 10.1038/sj.ijo.0802051] [Medline: 12187404]
45. Wing RR, Tate DF, Gorin AA, Raynor HA, Fava JL. A self-regulation program for maintenance of weight loss. N Engl J Med 2006 Oct 12;355(15):1563-1571. [doi: 10.1056/NEJMoa061883] [Medline: 17035649]
46. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 7;285(9):1172-1177. [Medline: 11231746]
47. 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. [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]
48. 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. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
49. 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. [doi: 10.1038/oby.2008.19] [Medline: 18309301]
50. Svetkey LP, Stevens VJ, Brantley PJ, Appel LJ, Hollis JF, Loria CM, Weight Loss Maintenance Collaborative Research Group. Comparison of strategies for sustaining weight loss: the weight loss maintenance randomized controlled trial. JAMA 2008 Mar 12;299(10):1139-1148. [doi: 10.1001/jama.299.10.1139] [Medline: 18334689]
51. 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]
52. Booth AO, Nowson CA, Matters H. Evaluation of an interactive, Internet-based weight loss program: a pilot study. Health Educ Res 2008 Jun;23(3):371-381 [FREE Full text] [doi: 10.1093/her/cyn007] [Medline: 18349032]
53. 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]
54. Agras WS, Taylor CB, Feldman DE, Losch M, Burnett KF. Developing computer-assisted therapy for the treatment of obesity. Behavior Therapy 1990 Dec;21(1):99-109. [doi: 10.1016/S0005-7894(05)80191-1]
55. DeLucia JL, Kalodner CR, Horan JJ. The effect of two nutritional software programs used as adjuncts to the behavioral treatment of obesity. J Subst Abuse 1988;1(2):203-208. [Medline: 2980870]
56. 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. [doi: 10.1038/oby.2009.85] [Medline: 19343018]
57. 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. [doi: 10.1038/oby.2007.584] [Medline: 17426316]
58. Wylie-Rosett J, Swencionis C, Ginsberg M, Cimino C, Wassertheil-Smoller S, Caban A, et al. Computerized weight loss intervention optimizes staff time: the clinical and cost results of a controlled clinical trial conducted in a managed care setting. J Am Diet Assoc 2001 Oct;101(10):1155-62; quiz 1163. [Medline: 11678486]
59. Shay LE, Seibert D, Watts D, Sbrocco T, Pagliara C. Adherence and weight loss outcomes associated with food-exercise diary preference in a military weight management program. Eat Behav 2009 Dec;10(4):220-227. [doi: 10.1016/j.eatbeh.2009.07.004] [Medline: 19778751]
60. Jones KS, Burkett LN. A Comparison of Instructional Methods for Weight Loss in Women. Health Promotion Practice 2002 Oct 01;3(4):501-506. [doi: 10.1177/152483902236720]

61. van Wier MF, Ariëns GA, Dekkers JC, Hendriksen IJ, Smid T, van Mechelen W. Phone and e-mail counselling are effective for weight management in an overweight working population: a randomized controlled trial. BMC Public Health 2009;9:6 [FREE Full text] [doi: 10.1186/1471-2458-9-6] [Medline: 19134171]
62. 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]
63. Hunter CM, Peterson AL, Alvarez LM, Poston WC, Brundige AR, Haddock CK, et al. Weight management using the internet a randomized controlled trial. Am J Prev Med 2008 Feb;34(2):119-126. [doi: 10.1016/j.amepre.2007.09.026] [Medline: 18201641]
64. Rothert K, Strecher VJ, Doyle LA, Caplan WM, Joyce JS, Jimison HB, et al. Web-based weight management programs in an integrated health care setting: a randomized, controlled trial. Obesity (Silver Spring) 2006 Feb;14(2):266-272. [doi: 10.1038/oby.2006.34] [Medline: 16571852]
65. Morgan PJ, Collins CE, Plotnikoff RC, Cook AT, Berthon B, Mitchell S, et al. Efficacy of a workplace-based weight loss program for overweight male shift workers: the Workplace POWER (Preventing Obesity Without Eating like a Rabbit) randomized controlled trial. Prev Med 2011 May;52(5):317-325. [doi: 10.1016/j.ypmed.2011.01.031] [Medline: 21300083]
66. Schroder KE. Computer-assisted dieting: effects of a randomised controlled intervention. Psychol Health 2010 Jun;25(5):519-534. [doi: 10.1080/08870440902812013] [Medline: 20204974]
67. Threlfall WE. Computer-assisted bibliotherapy for weight control. Dissertation Abstracts International 1984;45(6A):1697-1698.
68. Foster GD, Makris AP, Bailer BA. Behavioral treatment of obesity. Am J Clin Nutr 2005 Jul;82(1 Suppl):230S-235S [FREE Full text] [Medline: 16002827]
69. Cook RF, Billings DW, Hersch RK, Back AS, Hendrickson A. A field test of a web-based workplace health promotion program to improve dietary practices, reduce stress, and increase physical activity: randomized controlled trial. J Med Internet Res 2007;9(2):e17 [FREE Full text] [doi: 10.2196/jmir.9.2.e17] [Medline: 17581811]
70. Sternfeld B, Block C, Quesenberry CP, Block TJ, Husson G, Norris JC, et al. Improving diet and physical activity with ALIVE: a worksite randomized trial. Am J Prev Med 2009 Jun;36(6):475-483. [doi: 10.1016/j.amepre.2009.01.036] [Medline: 19460655]
71. Turnin MC, Bourgeois O, Cathelineau G, Leguerrier AM, Halimi S, Sandre-Banon D, et al. Multicenter randomized evaluation of a nutritional education software in obese patients. Diabetes Metab 2001 Apr;27(2 Pt 1):139-147 [FREE Full text] [Medline: 11353880]
72. Perri MG, Martin AD, Leermakers EA, Sears SF, Notelovitz M. Effects of group- versus home-based exercise in the treatment of obesity. J Consult Clin Psychol 1997 Apr;65(2):278-285. [Medline: 9086691]
73. Leermakers EA, Anglin K, Wing RR. Reducing postpartum weight retention through a correspondence intervention. Int J Obes Relat Metab Disord 1998 Nov;22(11):1103-1109. [Medline: 9822949]
74. 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]
75. Ahrens RA, Hower M, Best AM. Effects of weight reduction interventions by community pharmacists. J Am Pharm Assoc 2003;43(5):583-589. [Medline: 14626750]
76. Haapala I, Barengo NC, Biggs S, Surakka L, Manninen P. Weight loss by mobile phone: a 1-year effectiveness study. Public Health Nutr 2009 Dec;12(12):2382-2391. [doi: 10.1017/S1368980009005230] [Medline: 19323865]
77. Anderson ES, Winett RA, Wojcik JR, Winett SG, Bowden T. A computerized social cognitive intervention for nutrition behavior: direct and mediated effects on fat, fiber, fruits, and vegetables, self-efficacy, and outcome expectations among food shoppers. Ann Behav Med 2001;23(2):88-100. [Medline: 11394559]
78. Gabriele JM, Carpenter BD, Tate DF, Fisher EB. Directive and nondirective e-coach support for weight loss in overweight adults. Ann Behav Med 2011 Apr;41(2):252-263 [FREE Full text] [doi: 10.1007/s12160-010-9240-2] [Medline: 21108032]
79. 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]
80. 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]
81. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, Finnish Diabetes Prevention Study Group. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 2001 May 3;344(18):1343-1350. [doi: 10.1056/NEJM200105033441801] [Medline: 11333990]
82. Albarracín D, Gillette JC, Earl AN, Glasman LR, Durantini MR, Ho MH. A test of major assumptions about behavior change: a comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. Psychol Bull 2005 Nov;131(6):856-897 [FREE Full text] [doi: 10.1037/0033-2909.131.6.856] [Medline: 16351327]

83. de Bruin M, Viechtbauer W, Schaalma HP, Kok G, Abraham C, Hospers HJ. Standard care impact on effects of highly active antiretroviral therapy adherence interventions: A meta-analysis of randomized controlled trials. Arch Intern Med 2010 Feb 8;170(3):240-250. [doi: 10.1001/archinternmed.2009.536] [Medline: 20142568]
84. Abraham C, Graham-Rowe E. Are worksite interventions effective in increasing physical activity? A systematic review and meta-analysis. Health Psychology Review 2009 Mar;3(1):108-144. [doi: 10.1080/17437190903151096]
85. 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. [Medline: 18283207]

Abbreviations

BMI: body mass index
OQAQ: Overview Quality Assessment Questionnaire
RCT: randomized controlled trial
TTM: transtheoretical model

Edited by G Eysenbach; submitted 31.07.13; peer-reviewed by S Wieland, R Crutzen; comments to author 10.09.13; revised version received 04.12.13; accepted 21.12.13; published 19.02.14.

Please cite as:
Tang J, Abraham C, Greaves C, Yates T
Self-Directed Interventions to Promote Weight Loss: A Systematic Review of Reviews
J Med Internet Res 2014;16(2):e58
URL: http://www.jmir.org/2014/2/e58/
doi:10.2196/jmir.2857
PMID:24554464

Original Paper

Surviving and Thriving With Cancer Using a Web-Based Health Behavior Change Intervention: Randomized Controlled Trial

Erin O'Carroll Bantum[1], PhD; Cheryl L Albright[2], MPH, PhD; Kami K White[3], MPH; Jeffrey L Berenberg[4,5], MD; Gabriela Layi[1], MA; Phillip L Ritter[6], PhD; Diana Laurent[6], MPH; Katy Plant[7], MPH; Kate Lorig[6], DrPH

[1]University of Hawaii Cancer Center, Cancer Prevention & Control, Honolulu, HI, United States
[2]University of Hawaii at Manoa, Nursing and Dental Hygiene, Honolulu, HI, United States
[3]University of Hawaii Cancer Center, Biostatistics Shared Resource, Honolulu, HI, United States
[4]University of Hawaii Cancer Center, Clinical and Translational, Honolulu, HI, United States
[5]Tripler Army Medical Center, Medical Oncology, Honolulu, HI, United States
[6]Stanford University, School of Medicine, Stanford Patient Education Research Center, Palo Alto, CA, United States
[7]National Council on Aging, Washington, DC, United States

Corresponding Author:
Erin O'Carroll Bantum, PhD
University of Hawaii Cancer Center
Cancer Prevention & Control
701 Ilalo St, B4
Honolulu, HI, 96813
United States
Phone: 1 808 441 3491
Fax: 1 808 586 3077
Email: ebantum@cc.hawaii.edu

Abstract

Background: Given the substantial improvements in cancer screening and cancer treatment in the United States, millions of adult cancer survivors live for years following their initial cancer diagnosis and treatment. However, latent side effects can occur and some symptoms can be alleviated or managed effectively via changes in lifestyle behaviors.

Objective: The purpose of this study was to test the effectiveness of a six-week Web-based multiple health behavior change program for adult survivors.

Methods: Participants (n=352) were recruited from oncology clinics, a tumor registry, as well as through online mechanisms, such as Facebook and the Association of Cancer Online Resources (ACOR). Cancer survivors were eligible if they had completed their primary cancer treatment from 4 weeks to 5 years before enrollment. Participants were randomly assigned to the Web-based program or a delayed-treatment control condition.

Results: In total, 303 survivors completed the follow-up survey (six months after completion of the baseline survey) and participants in the Web-based intervention condition had significantly greater reductions in insomnia and greater increases in minutes per week of vigorous exercise and stretching compared to controls. There were no significant changes in fruit and vegetable consumption or other outcomes.

Conclusions: The Web-based intervention impacted insomnia and exercise; however, a majority of the sample met or exceeded national recommendations for health behaviors and were not suffering from depression or fatigue at baseline. Thus, the survivors were very healthy and well-adjusted upon entry and their ability to make substantial health behavior changes may have been limited. Future work is discussed, with emphasis placed on ways in which Web-based interventions can be more specifically analyzed for benefit, such as in regard to social networking.

Trial Registration: Clinicaltrials.gov NCT00962494; http://www.clinicaltrials.gov/ct2/show/NCT00962494 (Archived by WebCite at http://www.webcitation.org/6NIv8Dc6Q).

(J Med Internet Res 2014;16(2):e54) doi:10.2196/jmir.3020

KEYWORDS

cancer survivors; online interventions; social support

Introduction

In the United States, there are currently more than 12 million people who have survived cancer [1]. The rise in this number over recent years is an accomplishment and can be attributed, in large part, to better screening and treatment. However, physical and psychological sequelae may persist after treatment leading to chronic latent side effects, with survivors reporting symptoms that occur 12 months or longer post treatment [2,3]. Cancer-related fatigue is the most persistent side effect regardless of tumor or treatment type [4]. Fatigue and other side effects (eg, edema, pain) can also lead to depression and anxiety [5]. After recovering from months of cancer therapy or surgeries, many cancer survivors want to not only return to their previous lifestyle but often have an interest in making positive changes in their health and quality of life. The point in time where this interest occurs has been coined a “teachable moment” [6,7] and can serve as an opportune time to introduce health behavior change strategies regardless of the type of cancer, stage at diagnosis, or the presence of late effects.

Having good social support has been linked to better health outcomes and quality of life in cancer survivors [8-11]. Despite these developments, health behavior change interventions for cancer survivors are seldom conducted in a group setting where social support from other survivors is encouraged. Interventions that include a social support component (eg, support groups) are much more prevalent, especially when the intervention is focused on psychosocial behavior change (eg, anxiety and depression) [12,13]. Such psychosocial interventions have shown that interacting with other cancer survivors is strongly related to better adjustment in terms of mood and quality of life [14]; thus, such interactions might also facilitate changes in health behaviors.

Since more than 83% of adults aged 50-64 and 56% of adults aged 65 and over have access to high-speed Internet connections via computers, laptops, and smartphones [15], and due to the somewhat anonymous nature of online cancer survivors’ support groups, online venues have become appealing to adult cancer survivors. However, the impact of Web-based multiple health behavior change interventions tailored to cancer survivors has been limited. Some of the first online research was conducted by Gustafson and colleagues [16] and suggested that computer-based programs focused on specific physical or psychological symptoms could lead to improvements in those symptoms. Some of the initial work consisted largely of transferring the content of workbooks to an online format with very little interactivity or interaction between survivors [17]. Some Web-based distress management interventions have been associated with benefit in terms of mood, perceived stress, and cancer-related trauma [13,18], while other interventions have similar results for both face-to-face and online interventions with the same content (eg, sexual counseling following prostate cancer treatment) [19], with one study demonstrating a negative impact [20] on distress and quality of life.

Recent Web-based interventions designed to change health behaviors vary substantially in terms of their design, online features, and length of follow-up. Such differences in online features can make comparisons between the results from these interventions difficult. Some Web-based interventions have a social networking component [18] while other interventions serve as more of an online repository for information [21]. The length of interventions also vary greatly, with some interventions being short and structured [21], while others are much longer [22]. In addition to differing lengths of intervention trials for cancer survivors, most trials include participants with one type of cancer [13,18,21,23], while few trials have brought together people with a range of cancer types [22]. There has been a limited amount of research focusing on people with a range of cancer types or multiple health behaviors.

The Chronic Disease Self-Management Program (CDSMP) was developed for people with chronic conditions and focuses on multiple health behaviors. Because the population is heterogeneous, there is no expectation that all participants will make similar behavior changes. This program has been shown to be effective across numerous health conditions (eg, diabetes, arthritis) and across multiple formats (face-to-face groups and online groups) [24,25]. More detailed description is provided in the Methods section, but key components include: Action Planning, Problem Solving, Decision Making, and Self-Tailoring. CDSMP is facilitated by two trained peer facilitators, one or both of whom have experienced a chronic disease. Facilitators read every post or “comment” made by a course participant, stimulate peer-to-peer interactions, and personally advise participants about how to set realistic, confidence-building health behavior goals. This format allows for peer interaction as well as structured facilitation. In order to examine whether the CDSMP intervention would also be effective for cancer survivors, CDSMP was adapted for cancer survivors to create the “Surviving and Thriving with Cancer” (STC) intervention.

The STC trial tested the effectiveness of a tailored Web-based intervention to encourage multiple health behavior changes in post-treatment adult cancer survivors. In order to maintain consistency with the CDSMP, in addition to being variables of importance for cancer survivors, diet, exercise, depression, and fatigue were chosen as our outcomes of interest. We hypothesized that participants in the STC treatment condition would show six-month improvements in psychosocial symptoms including fatigue, insomnia, and depression, and would also report eating significantly more servings per day of fruits and vegetables when compared with participants in the wait-list control condition. We also hypothesized that participants in STC would report significantly more minutes of physical activity per week compared to controls.

Methods

Participants

Eligibility requirements for the STC trial were intentionally broad and included age (18 years of age or older), completion of primary treatment at least four weeks prior, but not more than 5 years before joining the study, diagnosis with only one cancer and no recurrence, access to the Internet, and ability to read English.

Recruitment

Potential participants were recruited via a number of different online and mailed strategies [26]. Recruitment at Stanford University was primarily conducted through online recruitment efforts, and in Hawaii, initial recruitment efforts focused on clinic-based recruitment in oncology offices on the island of Oahu, Tripler Army Medical Center (TAMC), and mailed recruitment letters to cancer survivors identified in the Tumor Registry at TAMC. However, in order to increase enrollment, recruitment methods were shifted to online nationwide recruitment via social networks used primarily by cancer survivors (Facebook, eg, “Throat Cancer Awareness”), Association of Cancer Online Resources (ACOR), and CURE print and digital magazine. Administrators of these various online sources were contacted and asked about willingness to post recruitment messages for their members. These messages invited interested people to learn more on the STC website. Approximately 60% (59.9%, 211/352) of the sample was recruited from these cancer-specific social media channels.

Study Design

We used a randomized controlled delayed-treatment design (NCT00962494). Participants were directed to the STC website and screened for eligibility and then completed an online consent approved by the University of Hawaii and Stanford Institutional Review Boards (IRB). Participants from TAMC completed a mailed consent form that was approved by the military IRB. Once consented, participants completed an online baseline questionnaire and were then randomized to treatment or control status. Randomization was conducted on a group-by-group basis. Once 40 to 50 participants had completed their baseline questionnaire, they were numbered in the order of completion and then randomized, using a random number table, half to treatment and half to wait-list control. All participants received a US $10.00 Amazon voucher for completing each questionnaire.

Intervention

The STC intervention was a six-week online workshop that was adapted from CDSMP [24], a patient education course adopting the underlying principle that people with similar health conditions can help each other improve their health behaviors. To create the STC program, a Web-based version of CDSMP was adapted to be more relevant for cancer survivors. The CDSMP’s modules on healthy eating were modified for cancer survivors living in Hawaii by adding foods that are commonly eaten in Hawaii, and modules on the changes in body, sleep, and other side effects associated with post-treatment recovery were added to the program.

Each cohort (group) consisted of approximately 20-25 survivors, with a total of nine cohorts. Each session of the six-week course included approximately 30-35 webpages of didactic material (in the “Learning Center” of the STC*)* that is geared towards skills building, information about specific content, and the encouragement of weekly action plans to build self-efficacy. Examples of content include improving diet by making healthier food choices, increasing exercise, stress management via relaxation training, improving communication with health care providers, processing and communicating emotional experiences to people inside of one’s existing social network, as well as group members, and fatigue management. More details of weekly topics can be found in Figure 1. At the end of each weekly educational session, users were invited to identify a health behavior they would like to change and were guided, in both the didactic materials, as well as by facilitators on how to set realistic, achievable goals, which were called action plans. These weekly action plans were posted on the “Discussion Center” (see below) and facilitators provided feedback and help. Participants were prompted both in the middle and at the end of a given week, via an automated message, to update the group on their progress as well as provide feedback to other group members.

Each group had two facilitators who were cancer survivors. The facilitators went through intensive online training about both the content of the intervention materials as well as how to respond to users’ comments and goals. They were mentored by the principal investigators, who during the course of the intervention also read all posts and gave feedback and help to the facilitators as needed.

The STC intervention website contained numerous unique components. The most crucial components were the “Discussion Center”, “My Tools”, “Post Office”, and “Help”. The Discussion Center feature of the website is where social networking occurred and survivors were encouraged to provide feedback and encouragement to each other. This was accomplished in four threaded bulletin boards: action planning, problem solving, difficult emotions, and celebrations. As discussed above, these were seeded from the materials in the Learning Center. In addition, participants could post directly to any of the four bulletin boards at any time. The My Tools component of the program allowed participants to use tools (eg, exercise logs) to help continue to shape their behavior on an individual basis. They could also listen to relaxation exercises and find links to resources outside of this intervention. The Post Office component allowed participants to message each other individually, including emailing the facilitators. While facilitators, mentors, and principal investigators had access to all posted messages, they were not specifically monitored as a way to ensure some level of confidentiality. In the Help component, participants could contact one of the website or study administrators for assistance, look over a tutorial of the website, and read the informed consent.

Figure 1. Topics included in the Surviving and Thriving with Cancer Intervention.

Workshop Overview

Topics	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6
Self-Management Overview	✓					✓
Stress/Symptom Management	✓		✓			✓
Solving Problems	✓			✓		
Finding Joy	✓					
Setting Goals	✓	✓	✓	✓	✓	✓
Making an Action Plan	✓	✓	✓	✓	✓	✓
Feedback		✓	✓	✓	✓	✓
Exercise		✓	✓			
Weight Management		✓				
Healthy Eating		✓		✓		
Body Changes		✓				
Difficult Emotions			✓		✓	
Communication			✓	✓		✓
Making Decisions			✓	✓		
Sleep			✓			
Fatigue			✓			
Future Plans				✓		✓
Effects of Treatment					✓	
Medications					✓	
Evaluating Treatments					✓	
Depression					✓	
Working with Your Health Care Team					✓	✓

Data Collection

Survey data were collected at two time points: baseline and six months later. Although it is typical to survey participants immediately after completion of the intervention, the goal in waiting was to see if any changes following the intervention were maintained. The delayed treatment control condition received no information or materials over this period.

Measures

Demographic and previous medical history items on the baseline questionnaire included: type and stage of cancer, date of diagnosis, course of treatment, co-morbidities, race/ethnicity, gender, marital status, and years of education. Measures were included to measure the following: fatigue, insomnia, exercise, fruit and vegetable intake, and depression. The Brief Fatigue Inventory (BFI) is a 15-item measure that was used to measure fatigue. It assesses both the severity of fatigue and the impact of fatigue on daily functioning during the last 24-hour period [27]. To measure insomnia, the 5-item validated Women's Health Initiative Insomnia Rating Scale (WHIIRS) [28] was used. This measures how often, on a 5-point scale (from "no, not in the past 4 weeks" to "yes, 5 or more times a week), the participant experiences trouble falling or staying asleep. The Godin Exercise Questionnaire was used to assess minutes per week of exercise in the categories of mild, moderate, and vigorous [29]. The Block Food Frequency Questionnaire [30] was used to identify how many fruits and vegetables were eaten in the previous week and the number of servings were counted to represent the total fruit and vegetable consumption. The Patient Health Questionnaire (PHQ-8) was used to measure depression. This 8-item measure asks individuals to rate how much, on a 4-point scale (with options ranging from “not at all” to “nearly every day”), a given DSM diagnostic criteria for depression is perceived [31].

Statistical Analysis

Baseline characteristics were reported as percentages for categorical variables and means and standard deviations for continuous variables. Differences between participants randomized to the control and intervention conditions were assessed using chi-square tests for categorical variables and *t* test for continuous variables. The primary analyses compared change from baseline to 6 months in the two conditions for the following outcome measures: fatigue, insomnia, minutes per week of physical activity (categorized as strenuous plus moderate aerobic, strenuous aerobic, moderate aerobic, mild aerobic, and stretching), servings of fruits and vegetables eaten per week, and depression. The physical activity outcome measures were transformed as (Y+1) to the 0.25 power, based on the Box-Cox method [31], to better meet model assumptions; all other outcomes were examined without transformation. Mixed linear models, including a random intercept term for each participant, were used to estimate and compare differences in outcomes over time between conditions. A second set of analyses was performed for the physical activity outcomes to address the many zero values reported by participants. A mixed-distribution model with random effects was used for these outcomes, simultaneously fitting a model for the probability of a value greater than zero and a model for the mean of values greater than zero [32]. The treatment effect was assessed by the *F* test of the fixed interaction parameter for time and intervention group. The effect size was computed by taking the differences between the means of the predicted values from the adjusted model at 6 months, divided by the standard deviation for the difference calculated from the within and between subject variance components. Models were adjusted for covariates selected a priori as likely to be related to the outcomes measures in this population. Adjustment variables included: age (continuous), race (white, nonwhite), gender, marital status (married, not married), smoking status (current, former, never), highest year of school completed (continuous), site of cancer diagnosis (breast, all others), cancer stage (in situ, stage 1, stage 2, stage 3, stage 4, unknown), and years since cancer was diagnosed. Subgroup analysis was performed by including a three-way interaction term between years since cancer diagnosis (≤2 or >2 years), condition group, and time, with all two-way interactions terms included. Model results are presented as means and 95% CIs of the predicted values obtained from the models.

Roughly 14% (13.9%, 49/352) of participants who were randomized did not provide any data at 6 months, which did not differ by condition (11.4%, 20/176 and 16.5%, 29/176) for control and intervention, respectively). To address attrition, correlates of attrition were identified using a logistic model regressing status (participants with data at 6 months vs participants with no 6-month data) onto baseline characteristics (same as adjustment variables listed above), condition group, and the presence of long term health conditions [including anxiety (yes, no), arthritis (yes, no), asthma (yes, no), back pain (yes, no), COPD (yes, no), depression (yes, no), diabetes (yes, no), high blood pressure (yes, no), heart disease (yes, no), sleep disorder (yes, no), and other (yes, no)], with a stepwise selection method.

Analyses were conducted using SAS, version 9.2. *P* values were two-sided and *P*<.05 was considered statistically significant.

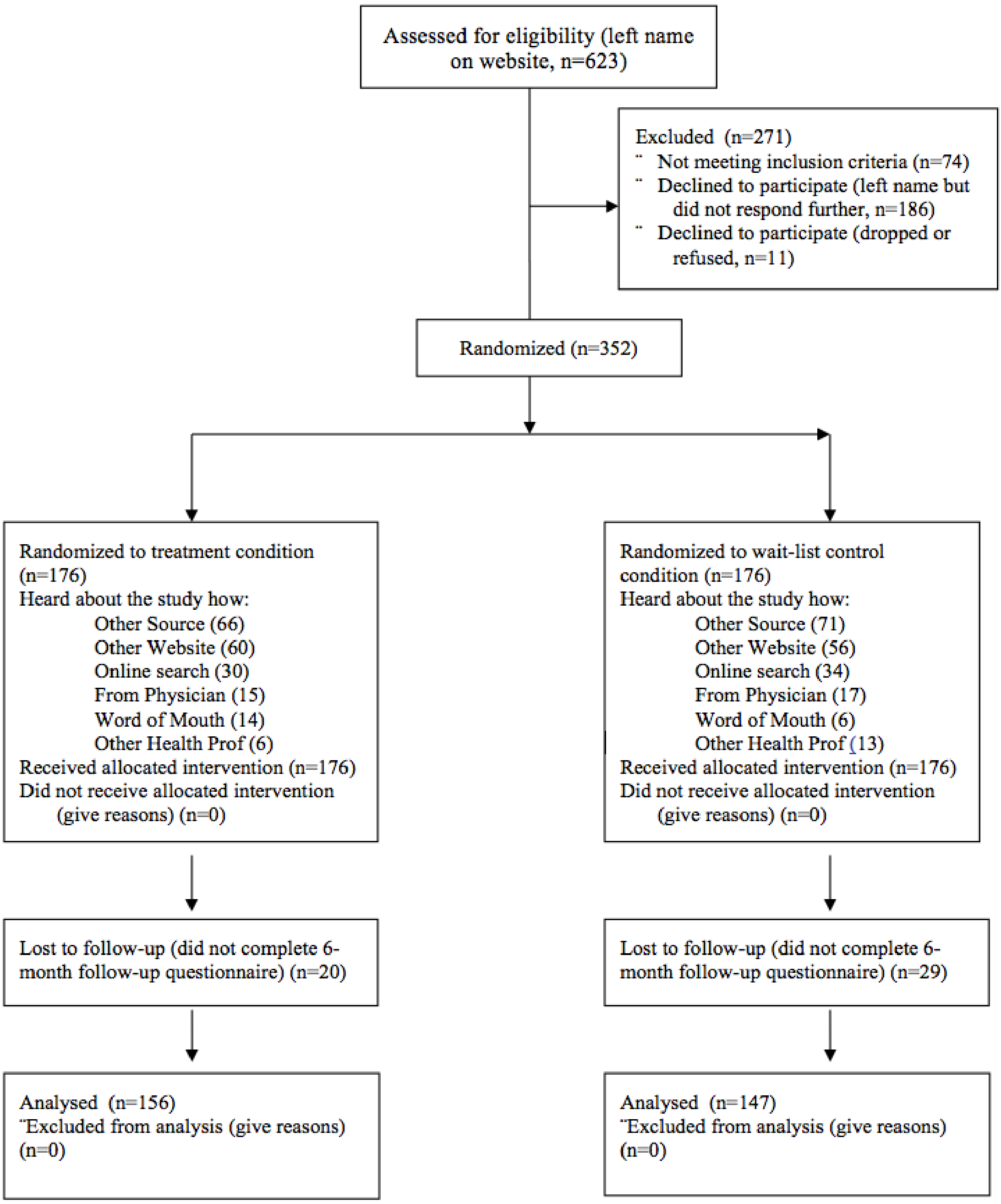
Results

Participants

Recruitment strategies are discussed in detail elsewhere [26]. Briefly, 60% (59.9%, 211/352) of the interested participants were recruited from online social networking sites, the rest were recruited from physician offices, a tumor registry attached to Tripler Army Medical Center, and a survivorship clinic on Oahu. Figure 2 provides the CONSORT recruitment diagram for the study (also see Multimedia Appendix 1 for details). Overall, 623 people were screened for eligibility, 352 people completed baseline measures, and 303 completed follow-up measures (n=156 in treatment condition; n=147 in control condition; see Figure 2 for details). In testing for predictors of dropout, less education (OR 0.84, 95% CI 0.75-0.95, per one year increase) and having long-term back pain (OR 2.31, 95% CI 1.13-4.75) was associated with dropout between baseline and 6 month follow-up.

The majority of participants were Caucasian (87.2%, 307/352) and female (82.1%, 289/352), having a mean age of 51 years (SD 11.2) and mean education level of 16 years (SD 2.9); 47.4% (167/352) were diagnosed with breast cancer and another 12.8% (45/352) of participants were given either an ovarian or uterine cancer diagnosis. Baseline characteristics of participants in the control and intervention groups are shown in Table 1. With the exception of age, no significant differences were found among the two groups. Additionally, there were no significant differences between the control and treatment groups on all outcomes measures at baseline. Participants in both groups reported mild levels of fatigue (mean scores of about 40 on the BFI), insomnia (mean scores of 9.6 on the WHIIRS), and were engaging in moderate plus strenuous activity with median values of 1.5 to 2 hours per week. Participants also reported eating, on average, 23 servings of fruit and vegetables each week.

Figure 2. CONSORT recruitment diagram.



Website Use

In regard to general use of the site, the mean number of sessions ever attended (logged on at least once) was 5.3 (SD 1.28) with the range being 0-6, and 67.0% (203/303) of participants attended all six sessions, with 86.8% (263/303) attending 4 or more sessions. There were 8016 total posts by treatment participants for an average of 46 posts per participant over the six-week intervention period.

Health Behavior Impact

Results for changes in health behaviors/psychosocial outcomes are reported in Table 2. Significant interactions between condition group and time were found for insomnia, strenuous exercise, and stretching exercise. The intervention group experienced an improvement from baseline to 6 months compared to the control group: reduced insomnia (9.6 to 9.2 compared to 9.6 to 10.1, *P*=.03), increased strenuous exercise (32 to 51 min/wk compared to a steady 29 min/wk, *P*=.01), and increased stretching (31 to 46 min/wk compared to 26 to 25 min/wk, *P*=.01). In the subgroup analyses looking at differences between survivors with diagnoses ≤ 2 and >2 years prior to enrollment, there were no significant differences, although there were suggested trends seen for both insomnia (*P*=.07) and depression (*P*=.09), such that people who were greater than 2 years post treatment improved slightly more on those measures (data not shown).

Table 1. Baseline characteristics of study population.

Characteristic		Control group, (n=176), n (%)	Intervention group, (n=176), n (%)	*P*
Age, mean (SD)		49.3 (11.0)	52.4 (11.0)	.008
Female		148 (84.1)	141 (80.1)	.33
Race				.84
	Caucasian	150 (85.2)	157 (89.2)	
	Asian	9 (5.1)	8 (4.6)	
	African American	4 (2.3)	2 (1.1)	
	American Indian/Alaskan Native	3 (1.7)	2 (1.1)	
	Native Hawaiian/Pacific Islander	3 (1.7)	1 (0.6)	
	Other	7 (4.0)	6 (3.4)	
Married		122 (69.3)	109 (61.9)	.14
Highest year education attained, mean (SD)		16.5 (3.1)	16.3 (2.8)	.62
Smoking status				.43
	Current	4 (2.3)	7 (4.0)	
	Former	57 (32.4)	64 (36.4)	
	Never	115 (65.3)	105 (59.7)	
Type of cancer diagnosed [a]				.70
	Breast	84 (47.7)	83 (47.2)	
	Endometrium/Uterine/Ovarian	23 (13.1)	22 (12.5)	
	Non-Hodgkins Lymphoma	13 (7.4)	7 (4.0)	
	Colorectal	11 (6.5)	11 (6.5)	
	Lung	7 (4.0)	8 (4.6)	
	Thyroid	6 (3.4)	8 (4.6)	
	Oral	6 (3.4)	5 (2.8)	
Stage of cancer				.96
	In situ	9 (5.1)	7 (4.0)	
	Stage 1	45 (25.6)	45 (25.6)	
	Stage 2	52 (29.6)	55 (31.3)	
	Stage 3	37 (21.0)	41 (23.3)	
	Stage 4	16 (9.1)	13 (7.4)	
	Unknown	17 (9.7)	15 (8.5)	
Number of years since cancer diagnosed, mean (SD)		2.5 (1.3)	2.4 (1.4)	.41
Number of years since treatment completed, mean (SD)		1.9 (1.2)	1.7 (1.2)	.09
Prevalence of long-term health conditions				
	High blood pressure	31 (17.6)	34 (19.3)	.68
	Depression	32 (18.2)	29 (16.5)	.67
	Back pain	25 (14.2)	29 (16.5)	.55
	Anxiety	29 (16.5)	23 (13.1)	.37
	Arthritis	18 (10.2)	27 (15.3)	.15
	Sleep disorder	18 (10.2)	12 (6.8)	.25
	Asthma	13 (7.4)	13 (7.4)	1.0

Characteristic		Control group, (n=176), n (%)	Intervention group, (n=176), n (%)	*P*
	Diabetes	9 (5.1)	11 (6.3)	.65
	Heart disease	3 (1.7)	8 (4.6)	.13
	Emphysema, COPD, chronic bronchitis	3 (1.7)	5 (2.8)	.72
	Other	39 (22.2)	39 (22.2)	1.0

[a] Sites also reported were oral cavity (n=11), soft tissue (n=11), testicular (n=10), kidney and renal (n=10), and other [n=26, including brain (n=5), prostate (n=4), eye (n=3)].

Table 2. Mean (95% CI)[a] of outcome measures from baseline to 6 months by condition group.

Outcome measures	Control group, mean (95% CI)		Intervention group, mean (95% CI)		*P* [b]	Effect size Month 6[c]
	Baseline (n=176)	Month 6 (n=156)	Baseline (n=176)	Month 6 (n=147)		
Fatigue (BFI[d])	40.8 (38.9-42.8)	40.7 (38.7-42.8)	39.0 (37.0-40.9)	36.4 (34.2-38.5)	.56	.17
Insomnia (WHIIRS[e])	9.6 (9.1-10.1)	10.1 (9.6-10.7)	9.6 (9.1-10.1)	9.2 (8.7-9.8)	.03	.20
Depression (PHQ[f])	7.7 (7.0-8.3)	7.1 (6.4-7.7)	6.5 (5.9-7.1)	6.1 (5.4-6.7)	.69	.19
Fruit/vegetable intake, times/week	22.7 (21.4-24.1)	23.2 (21.7-24.7)	24.3 (23.1-25.6)	25.9 (24.6-27.3)	.24	.21
Strenuous or moderate aerobic exercise, min/week	86.0 (72.3-99.7)	96.2 (79.9-112)	106 (91.1-120)	137 (119-155)	.45	.29
Strenuous aerobic exercise, min/week	29.0 (22.5-35.5)	28.9 (21.8-36.0)	32.0 (25.5-38.5)	50.8 (40.7- 60.9)	.01	.36
Moderate aerobic exercise, min/week	37.0 (30.9-43.2)	45.3 (37.5-53.0)	49.0 (42.2-55.7)	54.1 (46.5- 61.7)	.49	.10
Mild aerobic exercise, min/week	58.9 (51.5-66.2)	65.0 (56.5-73.6)	56.1 (48.9-63.3)	74.1 (64.2-84.1)	.28	.10
Stretching min/week	25.9 (21.3-30.4)	24.7 (20.0-29.5)	30.5 (25.1-35.8)	45.7 (38.1-53.4)	.01	.12

[a] Adjusted for age, race, sex, marital status, smoking status, education, years since diagnosis, site of cancer diagnosis, cancer stage. For outcomes of fatigue, insomnia, depression, and fruit/vegetable intake, means and 95% CIs were computed on the predicted values from the model. For outcomes of physical activity, means and 95% CIs were computed on the back-transformed predicted values (Y^4-1), where Y represented the predicted values from the model.

[b] Treatment effect was assessed by the *F* test of the fixed 2-way interaction parameter for time and condition group.

[c] Calculated by taking the differences of the means at 6 months predicted from the model, including adjustment factors, divided by the standard deviation for the difference computed from the within and between subject variance components.

[d] BFI: Brief Fatigue Inventory

[e] WHIIRS: Women's Health Initiative Insomnia Rating Scale

[f] PHQ: Patient Health Questionnaire

Discussion

Principal Findings

Participants in the treatment condition had significant reductions in insomnia and engaged in more strenuous and stretching exercises than those in the control condition. There is an established link between sleep disturbance and inflammation, which can be related to both cancer and depression [33], so impacting insomnia is a relevant finding. There is only one other known Web-based exercise and diet intervention for adult cancer survivors [34]. Although outcomes of that study are not yet available, we have found the current system usable and the intervention feasible. In regard to face-to-face interventions to impact exercise for cancer survivors, these have been demonstrated to be effective [35-37], often times with larger effect sizes than were demonstrated in this trial. This is crucial because Web-based interventions have relevance for people who have physical limitations or are not near facilities that could offer face-to-face interventions. Health behavior change interventions are relevant for cancer survivors, so continuing to test and refine interventions is imperative in the area of cancer survivorship.

Limitations

There are some limitations of the current study that should be noted. We measured health behaviors via self-report and there may have been over/underestimations of the dietary intake of fruits and vegetables, as well as physical activity, due to social desirability or recall bias. Due to significant economic, logistical, and noncompliance issues that can occur when nationwide online studies use objective measures for physical activity (eg, accelerometer) or telephone interviews for dietary intake (eg, 24-hour recall), this study was not able to use these types of assessments. That being said, self-reported health behaviors are commonly used for both Web-based and face-to-face trials and for several national health risk behavior surveys conducted by the NIH and CDC. Although the study focused on multiple outcomes, we did not adjust the significance level for multiple comparisons due to the exploratory nature of the analyses.

Our sample was well-educated and because more than half were recruited from various Internet sites, they had high levels of computer literacy and, thus, might be more familiar with posting their personal experiences on bulletin boards so others could comment on their success or lack thereof. Participants were not recruited or screened for entry based on specific inclusion or exclusion criteria for any specific health behavior (eg, low levels of physical activity or high levels of fatigue as criteria for eligibility) or for their inherent motivation/need to change all of the health behaviors addressed in the intervention. While this could have resulted in recruiting persons who were the most interested or more in need of changing a specific health behavior, in our study, it resulted in participants who were healthy, very well-adjusted, with little to no need (according to current recommendations) for significant changes in their health behaviors. At enrollment, their exercise and eating behaviors (in regard to intake of fruits and vegetables) were better than seen in national surveys, given that only 59% of average Americans eat the recommended 2.5 servings of vegetables per day and 42% eat the recommended 2.5 servings of fruits per day [38]. Scores on our depression measure indicate that participants, as a whole, had no concerns with depression. This leads to the question of whether we recruited survivors who were in need of support to improve multiple unhealthy habits, as well as whether this is one of the primary reasons that significant changes were not seen on many of the outcomes of interest (fatigue, depression, and increases in fruit and vegetable intake). The participants could also choose the behavior they wanted to change, regardless of their baseline level of that behavior or “need” to improve it. These factors could have contributed to the lack of significant change over the six-month period on some of the other outcome measures. In addition, when doing a population study where people enter with different concerns and a large range of scores on baseline measures, effect sizes can be muted. With a larger sample size, sensitivity analyses including only people who were not engaging in the health behaviors of interest at baseline could be explored. Future research could take into consideration these issues.

Another potential limitation is in regard to the lack of participants with a range of cancer types. As has been the case in the past and was the case with our study, the sample included a large percentage of female breast cancer survivors (47% of the sample), suggesting that the sample was more homogenous and perhaps the findings are less generalizable to people with other types of cancer. Future efforts for this to be more balanced are important and will be made in upcoming work. Although efforts were made to recruit people who would be more representative of cancer survivors as a whole in regards to gender, ethnicity, and cancer type, those efforts fell short in this study and continued efforts will be made.

Conclusions

Web-based interventions provide the ability to more fully understand the intervention aspects that are of most interest to cancer survivors, and with many of these interventions including social networking features, to understand the ways in which people interact and how that might be related to outcomes. People who have survived cancer clearly valued the social networking aspects of the STC site. There were multiple social networking components, such as webmail and numerous different discussion boards, so additional analyses could be conducted to understand what might be most important to the participants in terms of social networking. Understanding more about who people interacted with, as well as the content of those interactions, provides a foundation to more fully understand the ways in which people connect and how those connections matter in these sorts of interventions. Continued inclusion of social networking/online support in these types of interventions, as well as data collection on usage, is encouraged. Better understanding how the components included are used could also be a way to identify potent features of the intervention. It is important to note, though, that there could be synergistic effects that are difficult to capture technically when isolating components of interest. In conclusion, the Thriving and Surviving with Cancer intervention has been proven a relative success and additional efforts to understand what components are related to the most success could help further develop this, or any, Web-based intervention program.

Acknowledgments

There are many people to thank for their help and involvement with the study. We would first like to thank the participants for their willingness and interest in the study. We also thank Ross Yamato for his help with online recruitment and thank-you’s to participants. Thank you to Carolyn Gotay for securing the original funding for the project and to Ian Pagano for his help in some of the early statistical analyses, as well as Lynne Wilkens for her oversight of the analysis process and her helpful editing of the manuscript. We would also like to thank the Department of Defense and Stanford Cancer Center for funding this project (Department of Defense W81XWH-06-2-0042, Developmental Cancer Research Award from Stanford Cancer Center), in addition to all of the sites that helped with recruitment, including the facilitators of websites that posted our recruitment ad.

The views expressed in this manuscript are those of the authors and do not reflect the official policy or position of the Department of the Army, Department of Defense, or US Government.

Conflicts of Interest

KL published the book, Living a Healthy Life With Chronic Conditions, in 2006, which was given to the participants in this study as an intervention aid. KL receives royalties from this book but has no direct conflicts of interest with this study. All other authors have no conflicts of interest.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [39].

[PDF File (Adobe PDF File), 997KB - jmir_v16i2e54_app1.pdf]

References

1. Parry C, Kent EE, Mariotto AB, Alfano CM, Rowland JH. Cancer survivors: a booming population. Cancer Epidemiol Biomarkers Prev 2011 Oct;20(10):1996-2005 [FREE Full text] [doi: 10.1158/1055-9965.EPI-11-0729] [Medline: 21980007]
2. Boivin JF. Second cancers and other late side effects of cancer treatment. A review. Cancer 1990 Feb 1;65(3 Suppl):770-775. [Medline: 2406000]
3. Daniëls LA, Oerlemans S, Krol AD, van de Poll-Franse LV, Creutzberg CL. Persisting fatigue in Hodgkin lymphoma survivors: a systematic review. Ann Hematol 2013 Aug;92(8):1023-1032. [doi: 10.1007/s00277-013-1793-2] [Medline: 23728609]
4. Bower JE. Prevalence and causes of fatigue after cancer treatment: the next generation of research. J Clin Oncol 2005 Nov 20;23(33):8280-8282. [doi: 10.1200/JCO.2005.08.008] [Medline: 16219929]
5. Vahdaninia M, Omidvari S, Montazeri A. What do predict anxiety and depression in breast cancer patients? A follow-up study. Soc Psychiatry Psychiatr Epidemiol 2010 Mar;45(3):355-361. [doi: 10.1007/s00127-009-0068-7] [Medline: 19458878]
6. Demark-Wahnefried W, Aziz NM, Rowland JH, Pinto BM. Riding the crest of the teachable moment: promoting long-term health after the diagnosis of cancer. J Clin Oncol 2005 Aug 20;23(24):5814-5830 [FREE Full text] [doi: 10.1200/JCO.2005.01.230] [Medline: 16043830]
7. Lajous M, Mozaffarian D, Mozaffarian R, Schrag D, Adami HO. Lifestyle prescriptions for cancer survivors and their communities. J Intern Med 2011 Jan;269(1):88-93. [doi: 10.1111/j.1365-2796.2010.02273.x] [Medline: 21158981]
8. Lutgendorf SK, De Geest K, Bender D, Ahmed A, Goodheart MJ, Dahmoush L, et al. Social influences on clinical outcomes of patients with ovarian cancer. J Clin Oncol 2012 Aug 10;30(23):2885-2890 [FREE Full text] [doi: 10.1200/JCO.2011.39.4411] [Medline: 22802321]
9. Helgeson VS, Snyder P, Seltman H. Psychological and physical adjustment to breast cancer over 4 years: identifying distinct trajectories of change. Health Psychol 2004 Jan;23(1):3-15. [doi: 10.1037/0278-6133.23.1.3] [Medline: 14756598]
10. Hoyt MA, Stanton AL. Unmitigated agency, social support, and psychological adjustment in men with cancer. J Pers 2011 Apr;79(2):259-276. [doi: 10.1111/j.1467-6494.2010.00675.x] [Medline: 21395588]
11. Roland KB, Rodriguez JL, Patterson JR, Trivers KF. A literature review of the social and psychological needs of ovarian cancer survivors. Psychooncology 2013 Nov;22(11):2408-2418. [doi: 10.1002/pon.3322] [Medline: 23760742]
12. Classen CC, Kraemer HC, Blasey C, Giese-Davis J, Koopman C, Palesh OG, et al. Supportive-expressive group therapy for primary breast cancer patients: a randomized prospective multicenter trial. Psychooncology 2008 May;17(5):438-447 [FREE Full text] [doi: 10.1002/pon.1280] [Medline: 17935144]
13. 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]
14. Stanton AL, Revenson TA, Tennen H. Health psychology: psychological adjustment to chronic disease. Annu Rev Psychol 2007;58:565-592. [doi: 10.1146/annurev.psych.58.110405.085615] [Medline: 16930096]
15. Pew Internet.: Pew Research Center and American Life Project Spring Tracking Survey, April 17-May 19, 2013 URL: http://www.pewinternet.org/Static-Pages/Trend-Data-(Adults)/Whos-Online.aspx [accessed 2014-02-07] [WebCite Cache ID 6NDZU8gYg]
16. Gustafson DH, Hawkins RP, Boberg EW, McTavish F, Owens B, Wise M, et al. CHESS: 10 years of research and development in consumer health informatics for broad populations, including the underserved. Int J Med Inform 2002 Nov 12;65(3):169-177. [Medline: 12414016]
17. Gustafson DH, Bosworth K. CHESS: providing decision support for reducing health risk behavior and improving access to health services. Interfaces 1991;21(3):93-104.

18. 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]
19. Schover LR, Canada AL, Yuan Y, Sui D, Neese L, Jenkins R, et al. A randomized trial of internet-based versus traditional sexual counseling for couples after localized prostate cancer treatment. Cancer 2012 Jan 15;118(2):500-509 [FREE Full text] [doi: 10.1002/cncr.26308] [Medline: 21953578]
20. Salzer MS, Palmer SC, Kaplan K, Brusilovskiy E, Ten Have T, Hampshire M, et al. A randomized, controlled study of Internet peer-to-peer interactions among women newly diagnosed with breast cancer. Psychooncology 2010 Apr;19(4):441-446. [doi: 10.1002/pon.1586] [Medline: 19484712]
21. Carpenter KM, Stoner SA, Schmitz K, McGregor BA, Doorenbos AZ. An online stress management workbook for breast cancer. J Behav Med 2012 Dec 2:1. [doi: 10.1007/s10865-012-9481-6] [Medline: 23212928]
22. Owen JE, Bantum EO, Criswell K, Bazzo J, Gorlick A, Stanton AL. Representativeness of two sampling procedures for an internet intervention targeting cancer-related distress: a comparison of convenience and registry samples. J Behav Med 2013 May 4:1-12. [doi: 10.1007/s10865-013-9509-6] [Medline: 23645145]
23. Lepore SJ, Buzaglo JS, Lieberman MA, Golant M, Davey A. Standard versus prosocial online support groups for distressed breast cancer survivors: a randomized controlled trial. BMC Cancer 2011;11:379-390 [FREE Full text] [doi: 10.1186/1471-2407-11-379] [Medline: 21867502]
24. Lorig KR, Sobel DS, Stewart AL, Brown BW, Bandura A, Ritter P, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. Med Care 1999 Jan;37(1):5-14. [Medline: 10413387]
25. Lorig KR, Ritter PL, Laurent DD, Plant K. Internet-based chronic disease self-management: a randomized trial. Med Care 2006 Nov;44(11):964-971. [doi: 10.1097/01.mlr.0000233678.80203.c1] [Medline: 17063127]
26. Layi G, Albright CA, Berenberg J, Plant K, Ritter P, Laurent D, et al. UH Cancer Center Hotline: Recruiting cancer survivors for an online health-behavior change intervention: are different strategies more beneficial? Hawaii Med J 2011 Oct;70(10):222-223 [FREE Full text] [Medline: 22162599]
27. Mendoza TR, Wang XS, Cleeland CS, Morrissey M, Johnson BA, Wendt JK, et al. The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. Cancer 1999 Mar 1;85(5):1186-1196. [Medline: 10091805]
28. Levine DW, Kripke DF, Kaplan RM, Lewis MA, Naughton MJ, Bowen DJ, et al. Reliability and validity of the Women's Health Initiative Insomnia Rating Scale. Psychol Assess 2003 Jun;15(2):137-148. [Medline: 12847774]
29. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 1985 Sep;10(3):141-146. [Medline: 4053261]
30. Block G, Coyle LM, Hartman AM, Scoppa SM. Revision of dietary analysis software for the Health Habits and History Questionnaire. Am J Epidemiol 1994 Jun 15;139(12):1190-1196. [Medline: 8209877]
31. Spitzer RL. Validation and Utility of a Self-report Version of PRIME-MD - The PHQ Primary Care Study. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737]
32. Snedecor GW, Cochran WG. Statistical methods, eighth edition. Ames, Iowa State: University Press; 1989.
33. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature 2008 Jul 24;454(7203):436-444. [doi: 10.1038/nature07205] [Medline: 18650914]
34. Lee MK, Park HA, Yun YH, Chang YJ. Development and formative evaluation of a web-based self-management exercise and diet intervention program with tailored motivation and action planning for cancer survivors. JMIR Res Protoc 2013;2(1):e11 [FREE Full text] [doi: 10.2196/resprot.2331] [Medline: 23612029]
35. Craft LL, Vaniterson EH, Helenowski IB, Rademaker AW, Courneya KS. Exercise effects on depressive symptoms in cancer survivors: a systematic review and meta-analysis. Cancer Epidemiol Biomarkers Prev 2012 Jan;21(1):3-19 [FREE Full text] [doi: 10.1158/1055-9965.EPI-11-0634] [Medline: 22068286]
36. Demark-Wahnefried W, Morey MC, Sloane R, Snyder DC, Miller PE, Hartman TJ, et al. Reach out to enhance wellness home-based diet-exercise intervention promotes reproducible and sustainable long-term improvements in health behaviors, body weight, and physical functioning in older, overweight/obese cancer survivors. J Clin Oncol 2012 Jul 1;30(19):2354-2361 [FREE Full text] [doi: 10.1200/JCO.2011.40.0895] [Medline: 22614994]
37. Courneya KS, Mackey JR, Bell GJ, Jones LW, Field CJ, Fairey AS. Randomized controlled trial of exercise training in postmenopausal breast cancer survivors: cardiopulmonary and quality of life outcomes. J Clin Oncol 2003 May 1;21(9):1660-1668. [doi: 10.1200/JCO.2003.04.093] [Medline: 12721239]
38. Héroux M, Janssen I, Lam M, Lee DC, Hebert JR, Sui X, et al. Dietary patterns and the risk of mortality: impact of cardiorespiratory fitness. Int J Epidemiol 2010 Feb;39(1):197-209 [FREE Full text] [doi: 10.1093/ije/dyp191] [Medline: 19380370]
39. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

ACOR: Association of Cancer Online Resources
BFI: Brief Fatigue Inventory
CDSMP: Chronic Disease Self-Management Program
PHQ: Patient Health Questionnaire
STC: Surviving and Thriving with Cancer
TAMC: Tripler Army Medical Center
WHIIRS: Women’s Health Initiative Insomnia Rating Scale

Edited by D Mohr; submitted 12.10.13; peer-reviewed by W Kuijpers, D Linares; comments to author 22.11.13; revised version received 02.01.14; accepted 08.01.14; published 24.02.14.

Please cite as:
O'Carroll Bantum E, Albright CL, White KK, Berenberg JL, Layi G, Ritter PL, Laurent D, Plant K, Lorig K
Surviving and Thriving With Cancer Using a Web-Based Health Behavior Change Intervention: Randomized Controlled Trial
J Med Internet Res 2014;16(2):e54
URL: http://www.jmir.org/2014/2/e54/
doi:10.2196/jmir.3020
PMID:24566820

Original Paper

Attrition and Adherence in a Web-Based Distress Management Program for Implantable Cardioverter Defibrillator Patients (WEBCARE): Randomized Controlled Trial

Mirela Habibović[1*], PhD; Pim Cuijpers[2*], PhD; Marco Alings[3], MD, PhD; Pepijn van der Voort[4], MD; Dominic Theuns[5], PhD; Leon Bouwels[6], MD, PhD; Jean-Paul Herrman[7], MD, PhD; Suzanne Valk[8], MD; Susanne Pedersen[1,5,9,10*], PhD

[1]Center of Research on Psychology in Somatic Diseases, Department of Medical and Clinical Psychology, Tilburg University, Tilburg, Netherlands
[2]Department of Clinical Psychology, Vrije University, Amsterdam, Netherlands
[3]Department of Cardiology, Amphia Hospital, Breda, Netherlands
[4]Department of Cardiology, Catharina Hospital, Eindhoven, Netherlands
[5]Department of Cardiology, Erasmus Medical Center, Rotterdam, Netherlands
[6]Department of Cardiology, Canisius-Wilhelmina Hospital, Nijmegen, Netherlands
[7]Department of Cardiology, Onze Lieve Vrouwe Gasthuis, Amsterdam, Netherlands
[8]Department of Cardiology, Vlietland Hospital, Schiedam, Netherlands
[9]Department of Psychology, University of Southern Denmark, Odense M, Denmark
[10]Department of Cardiology, Odense University Hospital, Odense, Denmark
[*]these authors contributed equally

Corresponding Author:
Susanne Pedersen, PhD
Department of Psychology
University of Southern Denmark
Campusvej 55
Odense M, 5230
Denmark
Phone: 45 6550 7992
Fax: 45 13 466 2067
Email: sspedersen@health.sdu.dk

Abstract

Background: WEB-Based Distress Management Program for Implantable CARdioverter defibrillator Patients (WEBCARE) is a Web-based randomized controlled trial, designed to improve psychological well-being in patients with an implantable cardioverter defibrillator (ICD). As in other Web-based trials, we encountered problems with attrition and adherence.

Objective: In the current study, we focus on the patient characteristics, reasons, and motivation of (1) completers, (2) those who quit the intervention, and (3) those who quit the intervention and the study in the treatment arm of WEBCARE.

Methods: Consecutive first-time ICD patients from six Dutch referral hospitals were approached for participation. After signing consent and filling in baseline measures, patients were randomized to either the WEBCARE group or the Usual Care group.

Results: The treatment arm of WEBCARE contained 146 patients. Of these 146, 34 (23.3%) completed the treatment, 88 (60.3%) dropped out of treatment but completed follow-up, and 24 (16.4%) dropped out of treatment and study. Results show no systematic differences in baseline demographic, clinical, or psychological characteristics between groups. A gradual increase in dropout was observed with 83.5% (122/146) completing the first lesson, while only 23.3% (34/146) eventually completed the whole treatment. Reasons most often given by patients for dropout were technical problems with the computer, time constraints, feeling fine, and not needing additional support.

Conclusions: Current findings underline the importance of focusing on adherence and dropout, as this remains a significant problem in behavioral Web-based trials. Examining possibilities to address barriers indicated by patients might enhance treatment engagement and improve patient outcomes.

Trial Registration: Clinicaltrials.gov: NCT00895700; http://www.clinicaltrials.gov/ct2/show/NCT00895700 (Archived by WebCite at http://www.webcitation.org/6NCop6Htz).

(J Med Internet Res 2014;16(2):e52) doi:10.2196/jmir.2809

KEYWORDS

implantable cardioverter defibrillator; Web-based interventions; adherence; dropout; attrition

Introduction

The implantable cardioverter defibrillator (ICD) is a cardiac device that is implanted with leads in and on the heart in patients for the primary and secondary prevention of sudden cardiac death due to life-threatening ventricular tachyarrhythmias [1,2]. In case of a ventricular arrhythmia, the ICD paces or delivers a low- or high-voltage electric shock to the heart muscle to terminate the arrhythmia and return the heart to a normal rhythm [1,2]. The experience of the ICD shock is often described by patients as "getting kicked in the chest by a big horse", although patients' experiences vary considerably [3,4].

A subgroup of ICD patients experiences psychological distress after ICD implantation, such as anxiety, depression, post-traumatic stress, and impaired quality of life [5,6]. However, distress in patients may not necessarily be attributed to the device and ICD therapy, but also to the underlying disease [7,8] and patients' pre-implantation psychological profile [9]. The importance of monitoring ICD patients for psychological distress is increasingly being recognized [10], likely due to risk for morbidity and mortality associated with patient distress [11,12] and due to preliminary evidence suggesting that distress is still undertreated in this population [13]. One possible explanation for this is the lack of well-designed, validated, randomized controlled trials (RCTs), leaving us with insufficient knowledge for the establishment of a sufficient evidence base to inform clinical practice [14]. Results of previous trials have been promising with respect to reducing distress, but the majority of these trials had a high dropout rate, jeopardizing the external validity of these studies [15,16].

In order to make psychological treatment for ICD patients more patient-tailored, which may reduce dropout, the use of online Web-based interventions has been advocated [15,17]. There is evidence showing that Web-based interventions are as effective as face-to-face interventions [18,19], and might be able to overcome barriers such as travel burden, time constraints, and reluctance to seek help, and thus reach underserved groups of patients [20,21]. However, adherence and attrition also remain major challenges in these trials [22,23]. To date, the reasons for dropout are not well understood and deserve attention in their own right, in order to increase the success and applicability of results of future Web-based intervention trials.

In the current study, we will describe the attrition and adherence issues that we encountered during the "WEB-based distress management program for implantable CARdioverter dEfibrillator" patients (WEBCARE) trial (NCT00895700). The trial design paper was published previously [24]. In brief, the aim of WEBCARE was to reduce anxiety and depression and improve the quality of life of patients with a first-time ICD implant through a 12-week Web-based intervention (called "Leef met je hart" or "Live with your heart"), using cognitive behavioral therapy as the mainstay of treatment combined with psycho-education related to the ICD and relaxation techniques. Specific objectives were to (1) examine whether (a) completers, (b) patients who dropped out of treatment but remained in the study (filled in follow-up questionnaires), and (c) patients who dropped out of the treatment and the study, differ systematically on baseline demographic, clinical, and psychological characteristics, and (2) present descriptive data on patients' reasons for dropping out.

Methods

Patient Population

Consecutively implanted ICD patients from six hospitals in the Netherlands (ie, Erasmus Medical Centre, Rotterdam; Amphia Hospital, Breda; Catharina Hospital, Eindhoven; Onze Lieve Vrouwe Gasthuis, Amsterdam; Canisius Wilhelmina Hospital, Nijmegen; Vlietland Hospital, Schiedam) were approached for study participation between April 2010 and February 2013. Patients were eligible for participation if they fulfilled the following inclusion criteria: first-time ICD implant, age 18-75 years, proficient in the Dutch language, and with Internet access and a sufficient level of Internet skills. Exclusion criteria were the following: life expectancy less than 1 year, history of psychiatric illness other than affective/anxiety disorders, or on the waiting list for heart transplantation.

Data Collection

Patients were approached by the ICD nurse or ICD technician prior to or briefly after ICD implantation. They were informed both verbally and in writing about the study. If the patient met the inclusion criteria and was willing to participate, informed consent was signed. Patients who could not decide at that time were approached again after ICD implant while still hospitalized. Prior to discharge from the hospital, consented patients were provided with the first set of questionnaires (baseline) and their medical records were accessed for information on their demographic and clinical variables. After completing the questionnaires, patients returned them in a self-addressed and pre-stamped envelope to Tilburg University, Netherlands, which served as the core lab for WEBCARE. If the questionnaires were not returned within two weeks, patients received up to 3 reminder phone calls. Patients who did not want to participate but who were willing to give access to information from their medical records also signed an informed consent form. The study was approved by the Medical Ethics Committee of all participating centers and was conducted in accordance with the Declaration of Helsinki. All patients provided written informed consent.

Design and Randomization

After receiving the baseline questionnaires and signed informed consent, participants were randomly assigned to either of two conditions: (1) the WEBCARE (WC) group, receiving questionnaires at baseline, 3 months, 6 months, and 12 months by mail, and getting access to the Web-based intervention for a time period of 12 weeks to complete 6 modules online [24] in addition to usual care, or (2) the Usual Care (UC) group, receiving only questionnaires by mail at all time points and usual care.

Patients were randomized using block randomization by computer, randomizing 20 patients per hospital, at each time point. Randomization lists were generated by an independent, blinded statistician and sealed by a research assistant. For the current analyses, we will only focus on patients who were randomized to the WC group.

Descriptive Data

Patients who signed the informed consent form but who decided to quit the intervention and/or the study prematurely were contacted by telephone 12 weeks after randomization and asked why they had decided to quit. This time interval was chosen in order to not interfere with possible intervention effects (patients were allowed to work at their own pace, some chose to finish the intervention within the first two weeks, while others decided to do the 6 lessons within the last two weeks. For that reason, it was clear at 12 weeks who had quit or finished the intervention). Hence, patients were contacted at the time that they should have received their 3-month follow-up and finished the 6-module online course.

Intervention: “Live with your Heart” (Leef met je Hart)

The intervention was based on the previously developed Web-based treatment “Alles Onder Controle” *(*Everything Under Control*)* [25] and was for the purpose of the WEBCARE trial adapted for ICD patients. The Alles Onder Controle treatment was developed for the healthy depressed population and has proven to be effective in reducing distress [25,26]. The Web-based course for ICD patients is a 12-week intervention of 6 online lessons addressing distress based on the cognitive behavioral model (problem-solving treatment). The first lesson focused on psycho-education with respect to living with an ICD (eg, what are “normal” adaptation problems post ICD implantation). In the second lesson, patients received homework assignments and were provided with therapist feedback (feedback was provided by master’s-level psychologists and was intended as minimal guidance to help patients get through the lessons—encouraging patients to continue with the lessons and giving guidance on how to address their problems according to problem-solving theory). In addition, patients received a relaxation training CD, which they were allowed to use throughout the intervention.

Patients were allowed to work at their own time and pace; however, if a lesson was not finished within two weeks, a reminder email was sent, with up to 3 reminders per lesson. Patients could proceed to the next lesson only when the previous one was finished and the homework assignment was sent to the therapist. If patients did not log in within the first two weeks, a reminder email was sent. Twelve weeks after receiving the log-in information, patients’ accounts were automatically closed.

Measures

Demographic and Clinical Measures

Information on demographic variables (ie, age, gender, working status, marital status, education level) was collected through purpose-designed questions in the questionnaires, while information on clinical variables (ie, left ventricular ejection fraction [LVEF], QRS-width [electrocardiogram reading], New York Heart Association functional class [NYHA-class], presence of heart failure, use of cardiac and psychotropic medication) were extracted from patients’ medical records at the time of implantation by the implanting electrophysiologist or research nurses at the participating centers. The Charlson Comorbidity Index [27] was calculated based on self-report data and information from patients’ medical records.

Anxiety

The Generalized Anxiety Disorder scale (GAD-7) was used to assess anxiety [28]. The GAD-7 is a 7-item self-report questionnaire assessing anxiety symptoms in the past two weeks (eg, “Feeling nervous, anxious, or on the edge”). The GAD-7 is a reliable measure, with a Cronbach alpha of .92 and an intraclass correlation of .83 [28]. The 7 items are rated on a 4-point Likert scale from 0 (not at all) to 3 (almost every day) (score range 0-21), with a higher score indicating increased anxiety symptoms.

Depression

The Patient Health Questionnaire (PHQ-9) is a 9-item self-report measure of depression (eg, “Having little interest or pleasure in doing things”) that taps into the 9 diagnostic criteria for DSM-IV depressive disorder [29]. The PHQ-9 can establish provisional depressive disorder diagnoses as well as grade depressive symptom severity. Items are evaluated on a 4-point Likert scale from 0 (not at all) to 3 (almost every day) (score range 0-27), with a higher score indicating more depressive symptoms [29]. The PHQ-9 has excellent reliability with a Cronbach alpha of .91 and good validity, and has previously been used in cardiac patients [30].

Type D (Distressed) Personality

Type D personality was assessed with the 14-item Type D scale (DS14) [31], which consists of two 7-item subscales measuring Negative Affectivity (eg, “I often feel unhappy”) and Social Inhibition (eg, “I am a “closed” kind of person”) [32]. Items are answered on a 5-point Likert scale ranging from 0 (false) to 4 (true), with total scores on both subscales ranging from 0 to 28. A standardized cut-off ≥10 on both subscales defines individuals with a Type D personality, as Item Response Theory has indicated this to be the most optimal cut-off [31,32]. Both subscales are internally consistent, with a Cronbach alpha of .88 for Negative Affectivity and .86 for Social Inhibition. The test-retest reliability for the two subscales over a 3-month period were r=0.72 and 0.82, respectively [33].

Optimism and Pessimism

The Life Orientation Test (LOT) was used to assess optimism and pessimism [34]. Optimism was measured using a sum score of items 1, 4, 5, and 11; while pessimism was assessed with the sum score of items 3, 8, 9, and 12. Items are answered on a 5-point Likert scale from 0 (very much disagree) to 4 (very much agree). The total score for the optimism and pessimism subscales ranges between 0 and 16, with a higher score indicating higher levels of the respective trait [34].

Statistical Analyses

Continuous variables were compared using the Student's *t* test, while discrete variables were compared using the chi-square test. Data are represented as percentages for nominal variables and mean (SDs) for continuous variables. To compare groups on psychological variables, ANOVAs (analysis of variance) were performed. If group differences were observed, the Tukey-Kramer post-hoc test for unequal group sizes was used to identify which groups differed significantly. Descriptive data were coded and analyzed using "frequencies". A *P*<.05 indicated statistical significance. All tests were two-tailed. Data were analyzed using SPSS Statistics 19.0 for Windows.

Results

Patient Characteristics

A detailed description of the patient recruitment for WEBCARE is displayed in Figure 1. A total of 1024 patients were approached for participation, 735 (71.78%) were excluded due to not meeting the inclusion criteria (n=492), refusing to participate (n=192), or not returning baseline measures (n=51). Eventually 289 patients were randomized to either the WC group (n=146) or the UC group (n=143).

Demographic, clinical, and psychological characteristics of patients who were randomized to the WC group are shown in Table 1. The mean age of the group was 58.23 (SD 9.87) and 120 (82.2%, 120/146) were male patients. In addition, 106 (72.6%, 106/146) had a higher educational level.

Figure 1. Flowchart of patient recruitment.

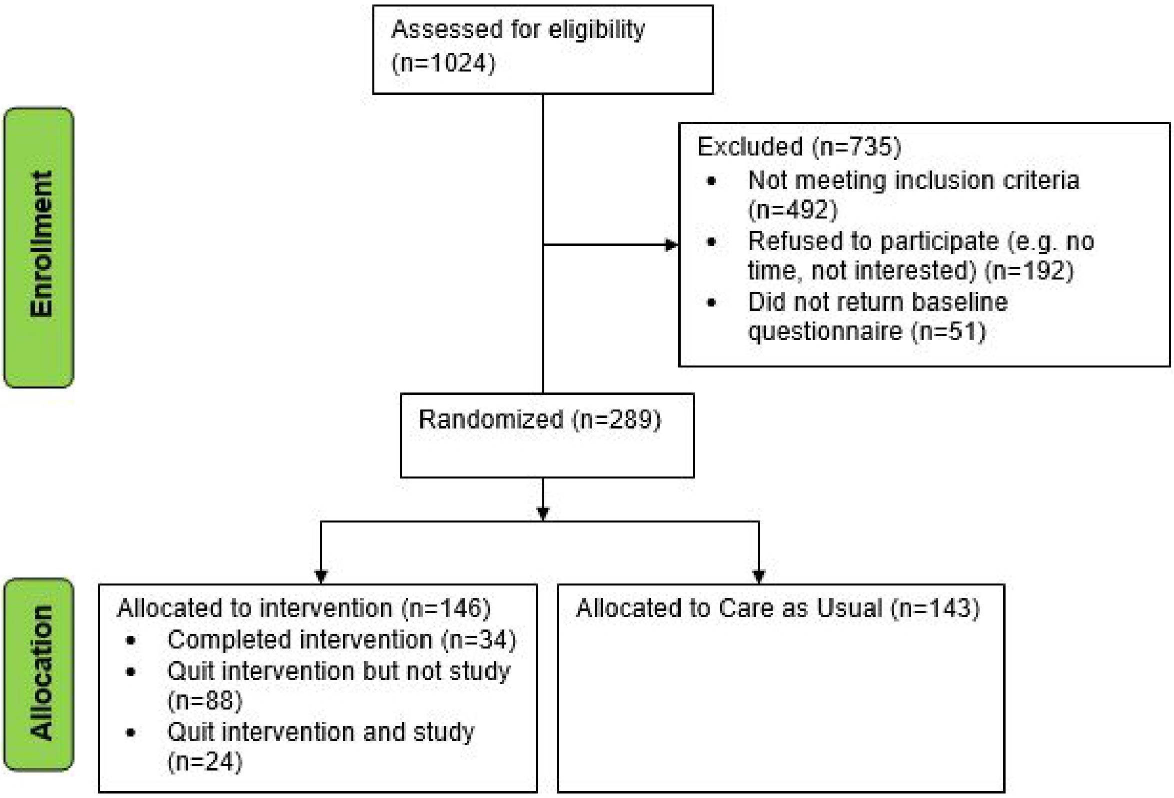


Table 1. Baseline demographic, clinical, and psychological characteristics of patients randomized to the WEBCARE treatment condition (WC; n=146).

Characteristic	WC group, mean (SD) or n (%)
Demographics	
Age, mean (SD)	58.23 (9.87)
Gender (male)	120 (82.2%)
Employed	68 (46.6%)
Has partner	124 (84.9%)
High education	106 (72.6%)
Clinical	
LVEF≤35[a]	87 (59.6%)
QRS>120ms[b], n=144	59 (41%)
NYHA class III/IV[c], n=122	20 (16.4%)
Heart failure	78 (53.4%)
CCI[d], mean (SD)	1.60 (1.06)
Beta-blockers	117 (80.1%)
ACE-inhibitors	82 (56.2%)
Diuretics	72 (49.3%)
Psychotropic medication	13 (8.9%)
Psychological	
Anxiety, mean (SD)	4.57 (5.02)
Depression, mean (SD)	5.93 (5.11)
Psychological treatment	8 (5.5%)
Cardiac rehabilitation, n=145	20 (13.8%)
Type D personality	24 (16.4%)
Optimism, mean (SD)	11.23 (2.68)
Pessimism, mean (SD)	5.73 (3.52)

[a] LVEF: left ventricular ejection fraction

[b] QRS-width: electrocardiogram Q,R, and S waves

[c] NYHA: New York Heart Association functional class

[d] CCI: Charlson Comorbidity Index

Non-Participants

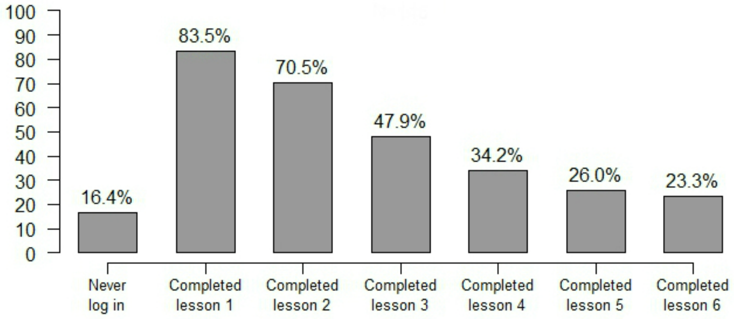
Patients who did not return the baseline questionnaires (n=51) and were excluded (not randomized) from current analyses did not differ systematically on demographic variables. However, significant differences on clinical variables were observed with patients who were not randomized, more often having a NYHA Class III/IV (*P*=.045), peripheral artery disease (*P*=.022), and using psychotropic medication (*P*≤.001) (anxiolytics, *P*=.004 and hypnotics, *P*=.010). Of the patients who refused to participate (n=192) but fulfilled the inclusion criteria, 60 (31.3%) signed consent and gave permission for medical record screening at the time of implantation. These patients were somewhat older [60.26 (SD 1.80) vs 58.16 (SD 10.30); *P*=.04], more likely to have a NYHA Class III/IV (34.1%, 14/41 vs 21.3%, 57/267; *P*=.013), to have experienced a previous myocardial infarction (76.6%, 36/47 vs 50.1%, 170/339; *P*=.001) or coronary artery bypass grafting (34.0%,16/47 vs 19.2%, 65/339; *P*=.019), to have peripheral artery disease (13.6%, 8/59 vs 5.3%, 18/339; *P*=.018), or to have a cardiac resynchronization therapy-defibrillator (CRT-D) (57.9%, 44/76 vs 25.9%, 87/336; *P*<.001) as compared to patients who signed informed consent for study participation (n=340).

Adherence and Attrition

Of the 146 randomized patients to the WC group, 34 (23.3%) completed the treatment and filled in the follow-up assessment (completers), 88 (60.0%) patients dropped out of the treatment but remained in the study and filled in the follow-up assessments (treatment dropouts), and 24 (16.4%) patients dropped out of the treatment and the study (dropouts). Focusing on the treatment, Figure 2 presents an overview of patients' adherence to the intervention and shows that the number of patients completing the lessons diminishes over time. The first lesson was completed by 83.5% (122/146) of the patients randomized to the WC group (16.5%, 24/146 never logged in), while only

23.3% (34/146) completed the last lesson (and thus the whole treatment schedule).

Figure 2. Adherence and attrition during the intervention (n=146).



Differences in Baseline Characteristics and Psychological Profile

Completers and (treatment) dropouts did not systematically differ on any baseline demographic, clinical, or psychological measures (Table 2). Although not significant, dropouts tended to be more often employed, have a lower education level, and lower mean score on anxiety and depression as compared to the other two groups. Table 2 shows a detailed overview of the baseline demographic and clinical characteristics and psychological profile of the three groups.

Table 2. Baseline demographic, clinical, and psychological characteristics stratified by group.

Characteristic	Completers, n=34	Treatment dropout, n=88	Dropout, n=24	*P*
	mean (SD) or n (%)			
Demographic				
Age, mean (SD)	57.91 (9.82)	58.84 (9.84)	56.42 (10.20)	.56
Gender (male)	28 (82.4%)	74 (84.1%)	18 (75.0%)	.59
Employed	15 (44.1%)	38 (43.2%)	15 (62.5%)	.23
Has partner	30 (88.2%)	73 (83.0%)	21 (87.5%)	.71
High education	26 (76.5%)	65 (74.7%)	15 (62.5%)	.43
Clinical				
LVEF≤35[a]	21 (61.8%)	52 (59.1%)	14 (58.3%)	.96
QRS≥120[b]	12 (36.4%)	37 (42.5%)	10 (41.7%)	.83
NYHA III/IV[c]	4 (13.3%)	13 (20.0%)	3 (15.0%)	.69
Heart failure	18 (52.9%)	46 (52.3%)	14 (58.3%)	.87
CCI[d], mean (SD)	1.88 (1.15)	1.56 (1.10)	1.38 (0.65)	.16
Beta-blockers	28 (82.4%)	69 (78.4%)	20 (83.3%)	.81
ACE-inhibitors	18 (52.9%)	52 (59.1%)	12 (50.0%)	.66
Diuretics	16 (47.1%)	43 (48.9%)	13 (54.2%)	.86
Psychotropics	1 (2.9%)	9 (10.2%)	3 (12.5%)	.36
Psychological				
Anxiety, mean (SD)	5.46 (5.18)	4.21 (5.08)	4.63 (4.62)	.47
Depression, mean (SD)	6.53 (4.40)	5.79 (5.38)	5.58 (5.17)	.73
Type D	7 (20.6%)	14 (15.9%)	3 (12.5%)	.70
Optimism, mean (SD)	11.12 (2.71)	11.41 (2.74)	10.75 (2.44)	.55
Pessimism, mean (SD)	6.15 (4.16)	5.56 (3.40)	5.79 (2.99)	.71

[a] LVEF: left ventricular ejection fraction

[b] QRS-width: electrocardiogram Q,R, and S waves

[c] NYHA: New York Heart Association functional class

[d] CCI: Charlson Comorbidity Index

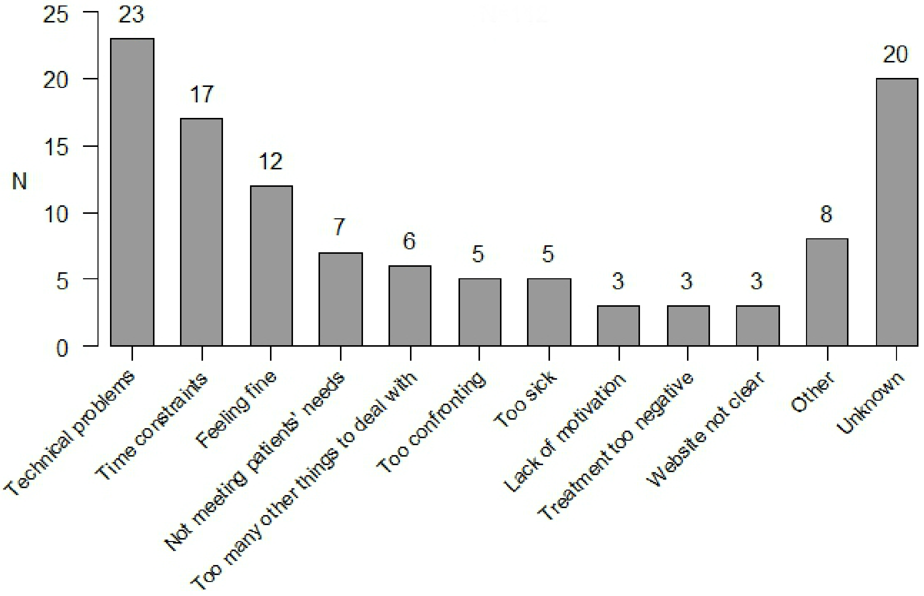
Descriptive Data: Reasons for Dropout

The reasons given by patients for not completing the treatment are displayed in Figure 3. Unfortunately, we were not able to reach all patients to learn about their reasons for dropout. Of the patients that we were able to reach, the majority (20.5%, 23/112) indicated that they faced technical problems (eg, the computer was not working, the website was not responding, problems with Internet connection): “My computer broke and I don’t use it that often, so I would have to buy another one just for the study; I didn’t wanted to do that” and “I got irritated because the website was not responding when I tried to log in, so I decided to quit”.

Time constraints were an issue in 15.2% (17/112) of patients (“I started working, so I don’t have time to do the online course”) and 10.7% (12/112) of patients were feeling fine and did not need additional support: “I had trouble thinking of any problems that I’m experiencing at this moment, as I don’t have any. I’m feeling fine”; and “I don’t have any problems. I’m happy that I have received the ICD. I feel reassured now”. Additional reasons were that the treatment did not apply to the patients’ needs (6.3%, 7/112): “I expected the course to be addressing only ICD specific problems…more technical problems. That’s why I decided to participate. I didn’t want to discuss psychological issues as I don’t have any”, and having already a lot to deal with (5.4%, 6/112): “I have just received an ICD and I’m dealing with depressive symptoms; I have other priorities at this moment”, and “I wanted to do the treatment but not within the 12 weeks. I wanted to start at a later time point because I had a lot to deal with immediately after the ICD implantation”.

Other reasons for dropout included treatment as being too confronting (4.5%, 5/112): “It was too personal, too confronting. I realized that I had more problems than I thought”, and feeling too sick (4.5%, 5/112): “It was too much. I had two surgeries in the past half year and I’m now on the waiting list for heart transplantation. I’m feeling sick all the time”. There were also patients who experienced the treatment as too negative: “I was feeling fine about my ICD, but when I started reading the content of the online course, I started feeling unhappy and therefore I decided to quit”, and “The homework assignments and questionnaires are too negatively worded, while you expect us to start thinking positive. I didn’t want to proceed as I didn’t wanted to start thinking negative about the ICD and how I’m feeling”.

Figure 3. Reasons for dropout (n=112).



Discussion

Principal Findings

This is the first behavioral randomized controlled trial to address the adherence and dropout issues of a Web-based intervention in the ICD population. Our findings show that 23.3% of patients randomized to the treatment arm completed the full treatment (six lessons), while 16.5% never logged on to the intervention. A gradual decline in adherence was observed with more patients dropping out as the lessons proceeded. The three groups (completers, treatment dropout, and dropout) did not differ systematically on any demographic or clinical baseline characteristics and their psychological profile. The top 3 reasons given for dropping out of the treatment were: technical issues with the computer/website, time constraints, and feeling fine/not needing additional support.

The findings of this study are generally in line with previous findings from the “Alles Onder Controle” Web-based intervention for individuals from the general population with increased levels of anxiety or depression who volunteered to undergo the intervention. Although generally higher percentages of completers were reported (38-55%), the rate of patients who never logged on was between 9% to 16% [25,26,35] and is in accordance with the 16.5%, which was observed in the current study. The higher number of completers in the Alles Onder Controle study might be attributed to these individuals being volunteers and thus highly motivated with respect to investing time and effort in improving their well-being. In addition, the latter patients scored high on anxiety and depressive symptoms, while patients in WEBCARE were included regardless of their anxiety/depression levels, as WEBCARE was set up as a prevention trial. However, there was a comparable gradual decline in adherence in the previous studies as in patients in the WEBCARE treatment arm [25,35].

In general, higher attrition rates are reported in open access Web-based trials without therapist guidance [36], while the number of completers is higher in closed trials where feedback is provided and reminders are sent [23,37]. Looking more specifically into cardiac patients (older adults with depression and comorbid cardiovascular disease), in their qualitative study on online treatment adherence, Donkin et al reported that time constraints, competing priorities, perception of limited worth of the intervention, and anxiety about spending time on the computer were important factors that contributed to an increased attrition [38], which are to some degree comparable to what we found in the current study. Generally, their results showed that adherence might increase if the benefits of participating in the treatment outweigh the costs of participation. In the current trial, the mean depression and anxiety scores were relatively low, patients were not distressed, and likely did not see a need for the treatment as they generally felt fine.

Studies of Web-based interventions have also shown that the duration of treatment [39] and timing [40] of the treatment may affect patients' engagement. Also, patients' perceived control over completing the course and coaching via phone vs email are found to be associated with higher completion rates [41]. In addition, printed delivery mode has shown to result in higher attrition as compared to a Web-based delivery mode [42]. These findings can partly explain why a higher attrition rate was observed in the current study. Current results show diminished participation after the third lesson. Perhaps, had the treatment been somewhat shorter, more patients would have completed all lessons. In addition, the treatment was offered immediately after ICD implantation. As indicated by some patients, at that time point they have a lot to deal with already and may for that reason decide not to participate or complete all modules.

Limitations and Strengths

A number of limitations of this study must be acknowledged. First, current analyses are based on a relatively small sample and should be replicated in larger studies in the future. Second, results on reasons for dropout are based on descriptive data that were obtained via a telephone call to patients. A structured interview or validated questionnaire would perhaps provide more valid information. Third, unfortunately we were not able to reach all patients at 12 weeks by telephone; hence, current findings are based on patients who answered the phone and were willing to provide us with information regarding their reasons for dropout.

This study also has several strengths. To our knowledge, it is the first behavioral intervention trial in ICD patients to have used a Web-based approach and the first study in ICD patients overall to address issues of adherence and attrition. The information from WEBCARE adds to our knowledge about factors that may influence adherence in trials using a Web-based approach, which can be used when designing Web-based behavioral intervention trials for ICD patients in the future, in order to increase the number of patients enrolled in the study and their treatment adherence.

When offering a Web-based intervention to ICD patients, it seems of great importance to make the intervention as patient-tailored as possible. Not all patients have the same needs at the same time. Thus, giving them time and space to complete the lessons when needed is an important factor as is making it possible to select which lessons to complete (some patients are more interested in technical aspects of the ICD, which would give them more reassurance, while others prefer psychological support in dealing with this new situation). As a proportion of the ICD patients indicated that they were feeling fine and did not need any additional support, it might be more important to focus on patients who have higher distress levels post implant. These patients could be identified using brief and standardized questionnaires that are designed to assess psychological distress. Close monitoring of patients' psychological needs is warranted as it has been associated with morbidity and mortality [11,12], and may help us to provide the right amount of care that is needed at the right time. Designing behavioral interventions with a collaborative care approach, where a period of "watchful waiting" is employed would perhaps be a way to go. With this approach, we would be able to offer support to patients who have "chronic" levels of distress or who develop distress at a later time point, as we have now learned that the "one size fits all" approach results in high dropout and low adherence. Incorporating a standard intervention in current health care models, which would be offered to all patients, might result in a great loss of resources as only a small proportion of patients would be willing to participate and complete the treatment (as shown in current data).

Conclusions

In conclusion, as Web-based treatments are increasingly being implemented in clinical practice, knowing how to keep patients motivated and compliant with treatment becomes more important every day. Our findings indicate that more attention should be paid to the technical aspects of Web-based treatment and making it more user-friendly. In addition, to overcome the barrier of home computers not working as they should, future studies should examine whether a similar intervention could be delivered using smartphones or tablets in order to decrease dropout. Also, future studies should examine the relationship between adherence and outcomes, as the results to date are inconclusive [22]. Examining the appropriate duration and timing of the intervention is also of great importance, which to date remains unexplored in the ICD population. The provision of patient-tailored interventions at the time when the patient needs it is likely to increase treatment adherence and enhance the effectiveness of such interventions.

Acknowledgments

This work was in part supported with grant no. 300020002 (with support of the Dutch Heart Foundation) and a VIDI grant (91710393) from the Netherlands Organization for Health Research and Development (ZonMW), The Hague, The Netherlands, to Dr Susanne S Pedersen. We would like to thank all the ICD and research nurses from the participating centers for their assistance with patient recruitment. In addition, we would like to thank Sophie Truijens, Harriët Abrahams, Leonie Visser, Eva Broers, Ferry van Ekelen, Annemiek de Wit, and Annick van Manen for their help with data management.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [43].

[PDF File (Adobe PDF File), 987KB - jmir_v16i2e52_app1.pdf]

References

1. Ezekowitz JA, Armstrong PW, McAlister FA. Implantable cardioverter defibrillators in primary and secondary prevention: a systematic review of randomized, controlled trials. Ann Intern Med 2003 Mar 18;138(6):445-452. [Medline: 12639076]
2. European Heart Rhythm Association, Heart Rhythm Society, Zipes DP, Camm AJ, Borggrefe M, Buxton AE, American College of Cardiology, American Heart Association Task Force, European Society of Cardiology Committee for Practice Guidelines. ACC/AHA/ESC 2006 guidelines for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: a report of the American College of Cardiology/American Heart Association Task Force and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Develop Guidelines for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death). J Am Coll Cardiol 2006 Sep 5;48(5):e247-e346. [doi: 10.1016/j.jacc.2006.07.010] [Medline: 16949478]
3. Sears SF, Kirian K. Shock and patient-centered outcomes research: is an ICD shock still a critical event? Pacing Clin Electrophysiol 2010 Dec;33(12):1437-1441. [doi: 10.1111/j.1540-8159.2010.02872.x] [Medline: 20727093]
4. Pedersen SS, Van Den Broek KC, Van Den Berg M, Theuns DA. Shock as a determinant of poor patient-centered outcomes in implantable cardioverter defibrillator patients: is there more to it than meets the eye? Pacing Clin Electrophysiol 2010 Dec;33(12):1430-1436. [doi: 10.1111/j.1540-8159.2010.02845.x] [Medline: 20663070]
5. Magyar-Russell G, Thombs BD, Cai JX, Baveja T, Kuhl EA, Singh PP, et al. The prevalence of anxiety and depression in adults with implantable cardioverter defibrillators: a systematic review. J Psychosom Res 2011 Oct;71(4):223-231. [doi: 10.1016/j.jpsychores.2011.02.014] [Medline: 21911099]
6. Habibović M, van den Broek KC, Alings M, Van der Voort PH, Denollet J. Posttraumatic stress 18 months following cardioverter defibrillator implantation: shocks, anxiety, and personality. Health Psychol 2012 Mar;31(2):186-193. [doi: 10.1037/a0024701] [Medline: 21806300]
7. Habibović M, Versteeg H, Pelle AJ, Theuns DA, Jordaens L, Pedersen SS. Poor health status and distress in cardiac patients: the role of device therapy vs. underlying heart disease. Europace 2013 Mar;15(3):355-361. [doi: 10.1093/europace/eus295] [Medline: 22989939]
8. Pedersen SS, Hoogwegt MT, Jordaens L, Theuns DA. Relation of symptomatic heart failure and psychological status to persistent depression in patients with implantable cardioverter-defibrillator. Am J Cardiol 2011 Jul 1;108(1):69-74. [doi: 10.1016/j.amjcard.2011.02.344] [Medline: 21529736]
9. Pedersen SS, Hoogwegt MT, Jordaens L, Theuns DA. Pre implantation psychological functioning preserved in majority of implantable cardioverter defibrillator patients 12 months post implantation. Int J Cardiol 2013 Jun 5;166(1):215-220. [doi: 10.1016/j.ijcard.2011.10.092] [Medline: 22071036]
10. Braunschweig F, Boriani G, Bauer A, Hatala R, Herrmann-Lingen C, Kautzner J, et al. Management of patients receiving implantable cardiac defibrillator shocks: recommendations for acute and long-term patient management. Europace 2010 Dec;12(12):1673-1690 [FREE Full text] [doi: 10.1093/europace/euq316] [Medline: 20974757]
11. Habibović M, Pedersen SS, van den Broek KC, Theuns DA, Jordaens L, van der Voort PH, et al. Anxiety and risk of ventricular arrhythmias or mortality in patients with an implantable cardioverter defibrillator. Psychosom Med 2013 Jan;75(1):36-41. [doi: 10.1097/PSY.0b013e3182769426] [Medline: 23197843]
12. Pedersen SS, Brouwers C, Versteeg H. Psychological vulnerability, ventricular tachyarrhythmias and mortality in implantable cardioverter defibrillator patients: is there a link? Expert Rev Med Devices 2012 Jul;9(4):377-388. [doi: 10.1586/erd.12.31] [Medline: 22905842]
13. Hoogwegt MT, Kupper N, Theuns DA, Zijlstra WP, Jordaens L, Pedersen SS. Undertreatment of anxiety and depression in patients with an implantable cardioverter-defibrillator: impact on health status. Health Psychol 2012 Nov;31(6):745-753. [doi: 10.1037/a0028018] [Medline: 22545981]

14. Habibović M, Burg MM, Pedersen SS. Behavioral interventions in patients with an implantable cardioverter defibrillator: lessons learned and where to go from here? Pacing Clin Electrophysiol 2013 May;36(5):578-590. [doi: 10.1111/pace.12108] [Medline: 23438053]
15. Pedersen SS, van den Broek KC, Sears SF. Psychological intervention following implantation of an implantable defibrillator: a review and future recommendations. Pacing Clin Electrophysiol 2007 Dec;30(12):1546-1554. [doi: 10.1111/j.1540-8159.2007.00905.x] [Medline: 18070312]
16. Salmoirago-Blotcher E, Ockene IS. Methodological limitations of psychosocial interventions in patients with an implantable cardioverter-defibrillator (ICD): A systematic review. BMC Cardiovasc Disord 2009;9:56 [FREE Full text] [doi: 10.1186/1471-2261-9-56] [Medline: 20040100]
17. Dickerson SS, Flaig DM, Kennedy MC. Therapeutic connection: help seeking on the Internet for persons with implantable cardioverter defibrillators. Heart Lung 2000;29(4):248-255. [doi: 10.1067/mhl.2000.108326] [Medline: 10900061]
18. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
19. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: A systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
20. Taylor CB, Jobson KO, Winzelberg A, Abascal L. The use of the Internet to provide evidence-based integrated treatment programs for mental health. Psychiat Ann 2002;32:671-677.
21. Proudfoot JG. Computer-based treatment for anxiety and depression: is it feasible? Is it effective? Neurosci Biobehav Rev 2004 May;28(3):353-363. [doi: 10.1016/j.neubiorev.2004.03.008] [Medline: 15225977]
22. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
23. 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]
24. Pedersen SS, Spek V, Theuns DA, Alings M, van der Voort P, Jordaens L, et al. Rationale and design of WEBCARE: a randomized, controlled, web-based behavioral intervention trial in cardioverter-defibrillator patients to reduce anxiety and device concerns and enhance quality of life. Trials 2009;10:120 [FREE Full text] [doi: 10.1186/1745-6215-10-120] [Medline: 20030843]
25. 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]
26. 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]
27. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [Medline: 3558716]
28. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
29. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
30. McManus D, Pipkin SS, Whooley MA. Screening for depression in patients with coronary heart disease (data from the Heart and Soul Study). Am J Cardiol 2005 Oct 15;96(8):1076-1081 [FREE Full text] [doi: 10.1016/j.amjcard.2005.06.037] [Medline: 16214441]
31. Denollet J, Sys SU, Stroobant N, Rombouts H, Gillebert TC, Brutsaert DL. Personality as independent predictor of long-term mortality in patients with coronary heart disease. Lancet 1996 Feb 17;347(8999):417-421. [Medline: 8618481]
32. Emons WH, Meijer RR, Denollet J. Negative affectivity and social inhibition in cardiovascular disease: evaluating type-D personality and its assessment using item response theory. J Psychosom Res 2007 Jul;63(1):27-39. [doi: 10.1016/j.jpsychores.2007.03.010] [Medline: 17586335]
33. Denollet J. DS14: standard assessment of negative affectivity, social inhibition, and Type D personality. Psychosom Med 2005;67(1):89-97 [FREE Full text] [doi: 10.1097/01.psy.0000149256.81953.49] [Medline: 15673629]
34. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. J Pers Soc Psychol 1994 Dec;67(6):1063-1078. [Medline: 7815302]
35. Warmerdam L, van Straten A, Jongsma J, Twisk J, Cuijpers P. Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: Exploring mechanisms of change. J Behav Ther Exp Psychiatry 2010 Mar;41(1):64-70. [doi: 10.1016/j.jbtep.2009.10.003] [Medline: 19913781]
36. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46 [FREE Full text] [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]

37. 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]
38. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
39. Christensen H, Griffiths KM, Mackinnon AJ, Brittliffe K. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. Psychol Med 2006 Dec;36(12):1737-1746. [doi: 10.1017/S0033291706008695] [Medline: 16938144]
40. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
41. Wojtowicz M, Day V, McGrath PJ. Predictors of participant retention in a guided online self-help program for university students: prospective cohort study. J Med Internet Res 2013;15(5):e96 [FREE Full text] [doi: 10.2196/jmir.2323] [Medline: 23697614]
42. Peels DA, Bolman C, Golsteijn RH, De Vries H, Mudde AN, van Stralen MM, et al. Differences in reach and attrition between Web-based and print-delivered tailored interventions among adults over 50 years of age: clustered randomized trial. J Med Internet Res 2012;14(6):e179 [FREE Full text] [doi: 10.2196/jmir.2229] [Medline: 23246790]
43. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

ANOVA: analysis of variance
CCI: Charlson Comorbidity Index
GAD: Generalized Anxiety Disorder scale
ICD: implantable cardioverter defibrillator
LVEF: left ventricular ejection fraction
NYHA class: New York Heart Association functional class
PHQ: Patient Health Questionnaire
QRS: electrocardiogram Q, R, and S waves
RCT: randomized controlled trial
UC: usual care (group)
WC: WEBCARE (group)
WEBCARE: WEB-based distress management program for implantable CARdioverter dEfibrillator patients

---

Edited by G Eysenbach; submitted 01.07.13; peer-reviewed by E Serber, G Boriani; comments to author 18.09.13; revised version received 11.10.13; accepted 10.12.13; published 28.02.14.

Please cite as:
Habibović M, Cuijpers P, Alings M, van der Voort P, Theuns D, Bouwels L, Herrman JP, Valk S, Pedersen S
Attrition and Adherence in a Web-Based Distress Management Program for Implantable Cardioverter Defibrillator Patients (WEBCARE): Randomized Controlled Trial
J Med Internet Res 2014;16(2):e52
URL: http://www.jmir.org/2014/2/e52/
doi:10.2196/jmir.2809
PMID:24583632

Original Paper

Evaluation of the Accuracy of Smartphone Medical Calculation Apps

Rachel Bierbrier[1]; Vivian Lo[1], BA(Hons), MASc; Robert C Wu[1,2], MSc, MD, FRCPC

[1]University Health Network, Centre for Innovation in Complex Care, Toronto, ON, Canada
[2]University Health Network, Department of General Internal Medicine, Toronto, ON, Canada

Corresponding Author:
Robert C Wu, MSc, MD, FRCPC
University Health Network
Department of General Internal Medicine
Toronto General Hospital, 14EN-222
200 Elizabeth Street
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 4567
Fax: 1 416 595 5826
Email: robert.wu@uhn.ca

Abstract

Background: Mobile phones with operating systems and capable of running applications (smartphones) are increasingly being used in clinical settings. Medical calculating applications are popular mhealth apps for smartphones. These include, for example, apps that calculate the severity or likelihood of disease-based clinical scoring systems, such as determining the severity of liver disease, the likelihood of having a pulmonary embolism, and risk stratification in acute coronary syndrome. However, the accuracy of these apps has not been assessed.

Objective: The objective of this study was to evaluate the accuracy of smartphone-based medical calculation apps.

Methods: A broad search on Google Play, BlackBerry World, and the iTunes App Store was conducted to find medical calculation apps for smartphones. The list of apps was narrowed down based on inclusion and exclusion criteria focusing on functions thought to be relevant by a panel of general internists (number of functions =13). Ten case values were inputted for each function and were compared to manual calculations. For each case, the correct answer was assigned a score of 1. A score for the 10 cases was calculated based on the accuracy of the results for each function on each app.

Results: We tested 14 apps and 13 functions for each app if that function was available. We conducted 10 cases for each function for a total of 1240 tests. Most functions tested on the apps were accurate in their results with an overall accuracy of 98.6% (17 errors in 1240 tests). In all, 6 of 14 (43%) apps had 100% accuracy. Although 11 of 13 (85%) functions had perfect accuracy, there were issues with 2 functions: the Child-Pugh scores and Model for End-Stage Liver Disease (MELD) scores on 8 apps. Approximately half of the errors were clinically significant resulting in a significant change in prognosis (8/17, 47%).

Conclusions: The results suggest that most medical calculating apps provide accurate and reliable results. The free apps that were 100% accurate and contained the most functions desired by internists were CliniCalc, Calculate by QxMD, and Medscape. When using medical calculating apps, the answers will likely be accurate; however, it is important to be careful when calculating MELD scores or Child-Pugh scores on some apps. Despite the few errors found, greater scrutiny is warranted to ensure full accuracy of smartphone medical calculator apps.

(J Med Internet Res 2014;16(2):e32) doi:10.2196/jmir.3062

KEYWORDS

cellular phone; mobile phone; mhealth; medical informatics applications; software; computers, handheld

Introduction

Smartphones are rapidly being adopted into the medical field. A recent survey found that 79% of medical students and 75% of postgraduate trainees owned smartphones [1]. One important use of smartphones is to aid in diagnosis, prognosis, and treatment of medical conditions. Apps can aid in diagnosis by providing a reference to staging systems, such as the severity staging of chronic obstructive pulmonary disease (COPD), or can provide rapid access to published algorithms in decision making. These reference or decision support functions that perform minimal calculations are typically considered to be at low risk of causing errors [2].

There are increasing numbers of clinical scoring systems that can include calculations, such as determining the severity of liver disease (Model for End-Stage Liver Disease, MELD), the likelihood of having a pulmonary embolism (Wells' Score for Pulmonary Embolism), and risk stratification in acute coronary syndrome (the thrombolysis in myocardial infarction, TIMI, score for non-ST elevation myocardial infarction, NSTEMI) [3-5]. Smartphone applications can make calculating these scores easier by providing information rapidly after performing a calculation using patient-specific data. Indeed, medical calculation apps are one of the most-used apps by doctors, often used several times per day [1]. Medical calculation apps can be considered to be of higher complexity because they do not just present previously published information, but may perform complex calculations based on user input. This increases the risk of error.

Health care professionals rely on decision-making aids such as medical apps, yet their accuracy has not been verified. The American Food and Drug Administration (FDA) has attempted to eliminate the distribution of faulty apps related to health care [6]. It is critical that apps used in clinical settings are accurate because the scoring results can impact a clinician's decision. Unfortunately, there is limited literature on the accuracy of smartphone medical calculators with the current evidence being highly specialized [7,8]. The purpose of our study was to assess the accuracy of general medical calculating apps on smartphones.

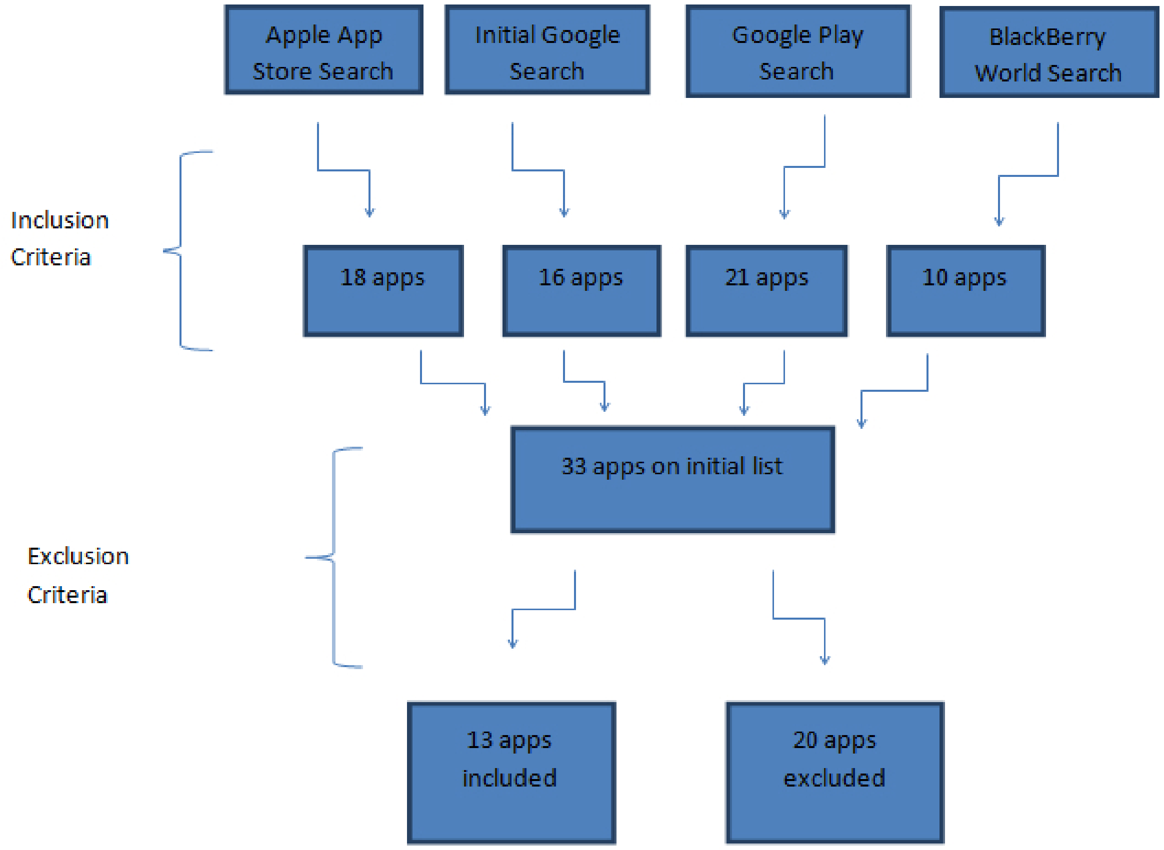
Methods

Definitions and Search Strategy

For the purpose of the study, an *app* was defined as a smartphone medical app. A *function* was defined as calculation that can be conducted on the app by inputting clinical data or observation. *General internists* were defined as specialists who apply scientific knowledge and clinical expertise to the diagnosis, treatment, and compassionate care of adults across the spectrum from health to complex illness [9].

Online searches were performed to acquire apps relevant to the study. The Google search contained the following keywords: “medical calculator apps,” “apps medical calculator,” “smartphone medical apps,” “medical + smartphone + apps,” and “medical + smartphone.” The first 5 pages of each of the searches were examined, with each page containing 10 links to websites. The keywords “medical” and “medical calculators” were then entered into the search fields of Google Play, BlackBerry World, and the iTunes App Store. The first 10 pages of each search on Google Play (24 apps per page) and BlackBerry World (6 apps per page) were examined. The first 30 rows (8 apps per row) of each search keyword were examined in the App Store for both iPhone and iPad. A complete breakdown of the app search can be found in Figure 1.

Figure 1. Breakdown of process to select apps for testing.



Inclusion and Exclusion Criteria

Two rounds of selection of apps occurred to acquire the final list of apps used for testing. The first round occurred as the apps were reviewed in the initial Google and App Store searches. Apps were included if they met the inclusion criteria of the study. Apps had to have a medical calculating smartphone app with 3 or more calculating functions.

Exclusion criteria were applied once the first draft of apps was compiled. This method of narrowing down the apps dealt with specific calculation functions of the apps focusing on apps in which a general internist would be interested. From all the smartphone apps from round 1, we compiled a list of all calculation functions. We provided this list of all calculation functions to 5 internists and asked them which functions they would want on a medical calculation application. Apps were excluded if it did not contain at least half of the functions selected by 5 physicians.

Testing

The medical apps were all downloaded in July 2013. To determine the functions to test, we used the preferred list of functions selected by the general internists. For functions to be tested, it had to be selected by at least 4 of the 5 internists. The selection process of functions is shown in Figure 2. Out of 476 calculating functions that were found on the apps, 147 (30.9%) were selected by 5 internists as useful functions they would want to have on an app. This list was then narrowed down further based on the degree of overlap to 15 functions (Figure 2). The Canadian Cardiovascular Society (CCS) Angina Score and the GOLD Classification of COPD were removed from the list of functions to test because they were classification systems without any calculations. A list of all calculation functions and descriptions is shown in Table 1.

Apps were tested on a single platform. Each function of each app was tested using the same 10 variations of data input, including 2 extremes and 8 middle values. The test cases were validated with clinicians for face validity. The aim of the different test cases was to produce variation in scores that would correspond to the different levels of severity that the functions contained. All the variations were recorded on an Excel spreadsheet. Answers from an app were considered correct if they were the same result as the calculation conducted using Excel with rounding error. All testing was conducted twice to reduce error. If an incorrect score was acquired, it was rechecked by another person. For each case, correct scores received a score of 1 and incorrect scores received a score of 0. Examples of calculator apps are shown in Figures 3-6.

Table 1. List of calculation functions.

Function	Description	# physicians choosing
$CHADS_2$	Scoring system for risk of stroke in atrial fibrillation (congestive heart failure, hypertension, age 75 years or older, diabetes mellitus, previous stroke or transient ischemic attack)	5
Child-Pugh score	Classification system for severity of liver disease	5
Wells' PE score	Scoring system for risk of pulmonary emboli (PE)	5
4T score	Scoring system for risk of heparin-induced thrombocytopenia (thrombocytopenia, timing, thrombosis, other)	4
ABCD2	Scoring system for risk of stroke after transient ischemic attack (TIA)-like symptoms (age, blood pressure, clinical features, duration of symptoms, and diabetes)	4
BMI	Body mass index	4
CIWA-Ar	Clinical Institute Withdrawal Assessment (CIWA) for Alcohol scale, revised	4
Corticosteroid conversion	Approximate equipotent dose conversions between different corticosteroids	4
HAS-BLED	Scoring system for risk of bleeding on anticoagulation (hypertension, abnormal renal/liver function, stroke, bleeding history or predisposition, labile international normalized ratio, elderly, drugs/alcohol concomitantly)	4
Creatinine clearance	Estimate of creatinine clearance by Cockcroft-Gault equation	4
MELD	Model for End-Stage Liver Disease (MELD) Scoring system for severity of liver disease, typically with United Network for Organ Sharing (UNOS) modifications	4
TIMI-STEMI	Thrombolysis in myocardial infarction (TIMI) risk stratification system after ST-elevation MI (STEMI)	4
TIMI-NSTEMI	TIMI risk stratification system after non–ST-elevation MI (NSTEMI)	4
CCS Angina Score[a]	Canadian Cardiovascular Society (CCS) Angina Score	4
GOLD classification[a]	GOLD classification of chronic obstructive pulmonary disease (COPD)	4

[a] Removed from the list of functions to test because is a classification system without any calculations.

Figure 2. Breakdown of process to select functions for testing.

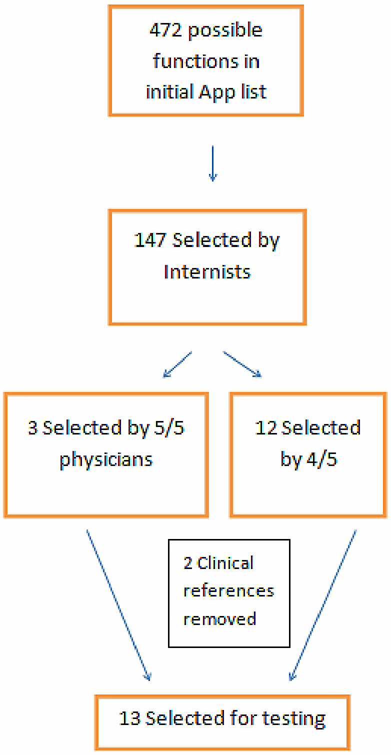


Figure 3. Calculations available on the Calculate by QxMD app.

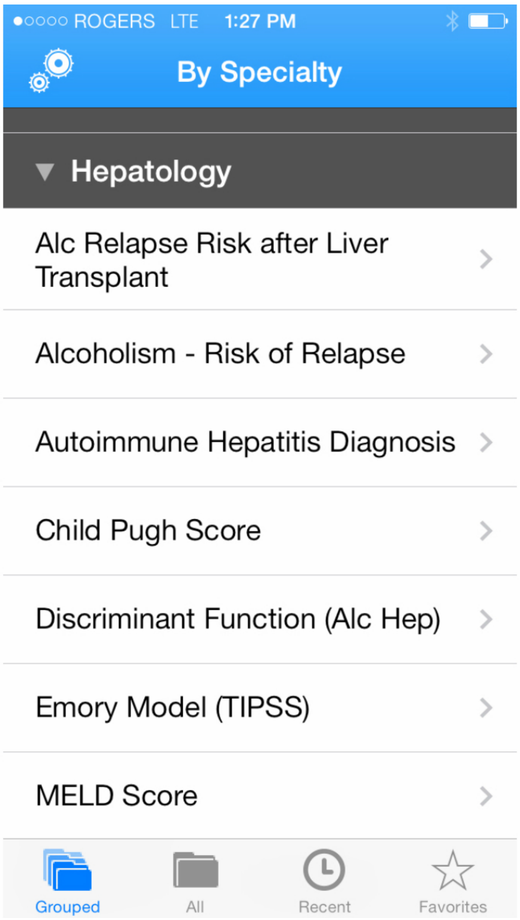
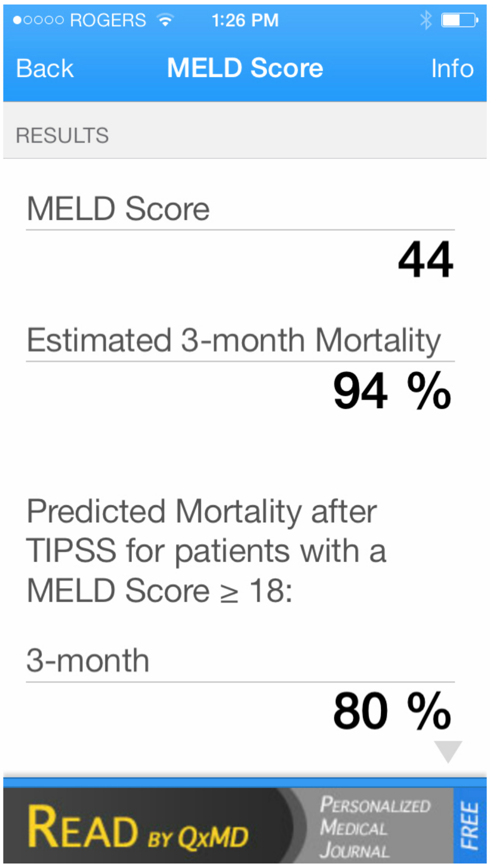


Figure 4. Example of a Model for End-Stage Liver Disease (MELD) Score calculation on the Calculate by QxMD app.



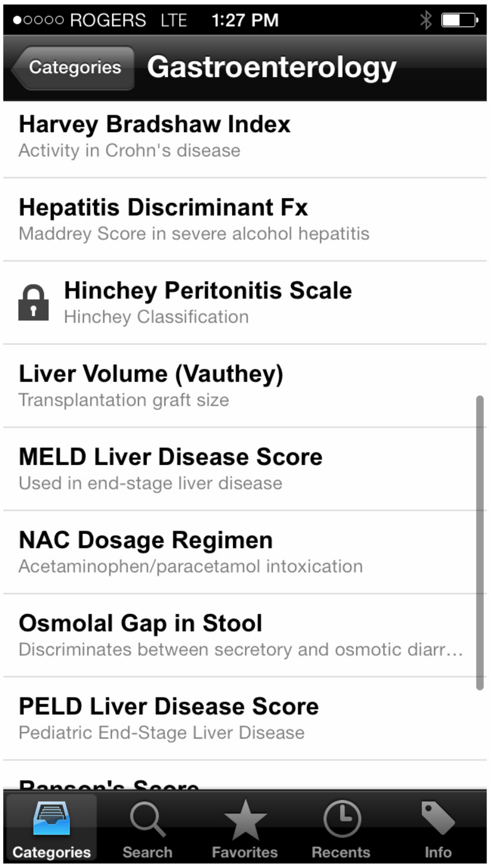
MELD Score
44

Estimated 3-month Mortality
94 %

Predicted Mortality after TIPSS for patients with a MELD Score ≥ 18:

3-month
80 %

Figure 5. Calculations available on the CliniCalc app by Medicon Apps.



Harvey Bradshaw Index
Activity in Crohn's disease

Hepatitis Discriminant Fx
Maddrey Score in severe alcohol hepatitis

Hinchey Peritonitis Scale
Hinchey Classification

Liver Volume (Vauthey)
Transplantation graft size

MELD Liver Disease Score
Used in end-stage liver disease

NAC Dosage Regimen
Acetaminophen/paracetamol intoxication

Osmolal Gap in Stool
Discriminates between secretory and osmotic diarr…

PELD Liver Disease Score
Pediatric End-Stage Liver Disease

Figure 6. Example of a Model for End-Stage Liver Disease (MELD) Score calculation on the CliniCalc app by Medicon Apps.

Results

The inclusion and exclusion criteria enabled us to come up with a list of frequently downloaded apps that were relevant to internal medicine. Fourteen (0.87%) smartphone apps were tested out of the 1603 smartphone apps found during initial research (Figure 1; Table 2).

Results of testing the 14 apps by using 10 variations for each of the 13 calculating functions are shown in Table 3. Out of the 1240 tests conducted there were 17 errors; therefore, the overall accuracy was 98.6% (17/1240).

In terms of functions, 11 of 13 functions (85%) were 100% accurate on all apps. The Child-Pugh score and the MELD score were 97% and 95% accurate, respectively. For the Child-Pugh score, there were errors in scoring for 2 apps. In all 4 errors found, the errors caused a difference in score by 1 point which did not translate to a different Child-Pugh class.

Issues occurred with the MELD score calculations on multiple apps. Eight of 14 apps produced similar incorrect scores for the cases involving creatinines >4 mg/dL (353.6 µmol/L). For 1 case, this error translated to an increased score which then gave an elevated severity (from 52.6% mortality to 71.3% mortality). This same error was found in 8 apps. These errors appeared to be because of incomplete application of United Network for Organ Sharing (UNOS) modifications of the original MELD scoring by the apps. The UNOS modification set a maximum allowable creatinine of 4 mg/dL (353.6 µmol/L).

Table 2. List of apps (accessed September 13, 2013).

App name	Developer(s)	Platforms available[a]	Platform tested[a]	Version tested	Cost ($US)
Calculate by QxMD [10]	QxMD	iOS, Android, BlackBerry	iOS	3.5.3	Free
CliniCalc [11]	Medicon Apps	iOS	iOS	2.1	Free
Epocrates [12]	epocrates	iOS, Android, BlackBerry	iOS	13.6	Free
MedCalc [13]	Pascal Pfiffner and Mathias Tschopp	iOS	iOS	2.7.3	$1.99
MedCalcs [14]	Beijing Kingyee Technology Co	iOS	iOS	2.6	Free
Medical Calculator [15]	Avivonet	Android	Android	1.0	$1.99
Medical Tools [16]	Irtza Sharif	Android	Android	1.2.1	Free
MediCalc [17]	ScyMed	iOS, Android	iOS	8.0	Free
MediMath [18]	Evan Schoenberg	iOS	iOS	4.3	$4.99
Mediquations [19]	Mediquations	iOS, Android	iOS	34.1	$4.99
MedScape [20]	WebMD, LLC	iOS, Android	iOS	4.2	Free
MedSolve Medical Calculator [21]	Charles Vu	iOS	iOS	1.2.2	$0.99
Skyscape Medical Resources [22]	Skyscape	iOS, Android, BlackBerry	iOS	1.18.42	Free
UpToDate [23]	UpToDate	iOS, Android, Windows	iOS*	1.3.7[b]	$563[c]

[a] iOS: iPhone/iPad/iPod operating system.

[b] The version tested was the online version on iOS platform.

[c] Requires subscription with MobileComplete and rates vary depending on role of user, country of user, and subscription term.

Table 3. Accuracy of medical calculating apps.

Name of app	$CHADS_2$	Child-Pugh	Wells' PE score	4T Score	ABCD2	BMI	CIWA-Ar	Corticosteroid conversion	HAS-BLED	Creatinine	MELD	TIMI-STEMI	TIMI-NSTEMI
Calculate by QxMD	100%	100%	100%	100%	100%	100%	—	100%	100%	100%	100%	100%	100%
CliniCalc	100%	100%	100%	100%	100%	100%	—	100%	100%	100%	100%	100%	100%
Epocrates	—	—	—	—	—	100%	—	100%	—	100%	80%	—	—
MedCalc	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	90%	100%	100%
MedCalcs	100%	100%	100%	—	100%	100%	—	—	100%	100%	80%	100%	100%
Medical Calculator	100%	90%	100%	—	100%	100%	—	—	—	100%	80%	100%	100%
Medical Tools	100%	100%	100%	—	—	100%	100%	100%	—	—	80%	—	100%
MediCalc	—	100%	—	—	—	100%	—	—	—	100%	100%	—	—
MediMath	100%	70%	100%	—	100%	100%	—	—	—	100%	80%	100%	100%
Mediquations	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
MedScape	100%	—	100%	—	—	100%	—	100%	—	100%	100%	100%	100%
MedSolve Medical Calculator	100%	100%	100%	—	—	100%	—	—	—	—	80%	100%	100%
Skyscape Medical Resources	100%	100%	100%	—	—	100%	—	—	—	100%	90%	—	100%
UpToDate	100%	100%	100%	—	—	100%	100%	100%	—	—	100%	—	100%

Discussion

The results of the study suggest that most medical calculator smartphone apps are accurate and can confidently be used in clinical settings. From an internal medicine perspective, the free apps that were 100% accurate and contained the most functions desired by internists were CliniCalc [11], Calculate by QxMD [10], and Medscape [20]. Although most of the apps provided accurate results, it is important to be cautious while using the Child-Pugh score and MELD score on certain apps, specifically.

There is a lack of evidence on the accuracy of medical calculating apps for smartphones. Information recommending medical calculating apps only provided qualitative information on the apps, without testing accuracy [24-27]. This study determines the actual accuracy of information provided by apps.

The study highlights the need for verifying medical apps before use in patient care. Although we found smartphone apps to be quite accurate, we found errors in the smartphone calculations that were clinically significant. There are efforts in the United States by the FDA to regulate medical device apps, but it is not clear if medical calculating apps are defined as medical devices in all countries [2,28]. Medical smartphone apps may be considered devices depending on the complexity of the patient information and calculation [2]. For medical apps that provide erroneous results, although downloaded from a global app store, they likely fall under legislation of the country where they are downloaded and used. The legal ramifications could be complex. Ultimately, it is likely the responsibility of the physician to determine if their calculating app is accurate. For individual physicians, testing and verifying each calculating function of each app is not reasonable. Thus, we provide physicians with clear evidence-based advice on which current apps to use.

Apps change quickly with new apps and frequent updates. We recommend that a system be put in place to verify smartphone apps that perform medical calculations to ensure they function properly. One way this can be done is by having a third party verify the accuracy of smartphone calculations. This could be conducted similar to our study but on a larger scale, with more variations and functions tested. With a list of trustworthy and validated apps, health care professionals could more confidently integrate smartphone technology into clinical settings.

There were limitations to this study. Because of the time frame of the project and the wide range of apps available, it was determined that not every function on every app could be tested. Thus, accuracy for each app may differ for other calculating functions or other test cases. Furthermore, although apps were available on multiple platforms, we only tested 1 platform for each app. Predominantly, the iOS platform was tested because most apps were available on this platform. Another limiting factor was that the focus was on apps used by general internists; therefore, results may differ for other specialties. However, this generalist approach does provide information on accuracy of smartphone medical calculation apps used by internists.

In summary, we found that most smartphone medical calculator app functions were accurate. However, some errors were noted in some functions of some apps. Given that using smartphones as medical calculators makes them a medical device, a system to verify smartphone calculation accuracy would be useful to reduce the chance of errors affecting patient care.

Acknowledgments

We would like to acknowledge the insightful observations by our colleague, Dr David K Wong, who led to the study conception.

Conflicts of Interest

None declared.

References

1. Payne KB, Wharrad H, Watts K. Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]
2. Sherwin-Smith J, Pritchard-Jones R. Medical applications: the future of regulation. Bull R Coll Surg Engl 2012 Jan 01;94(1):12-13 [FREE Full text] [doi: 10.1308/147363512X13189526438512]
3. Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PC. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. Hepatology 2000 Apr;31(4):864-871. [doi: 10.1053/he.2000.5852] [Medline: 10733541]
4. Antman EM, Cohen M, Bernink PJ, McCabe CH, Horacek T, Papuchis G, et al. The TIMI risk score for unstable angina/non-ST elevation MI: A method for prognostication and therapeutic decision making. JAMA 2000 Aug 16;284(7):835-842. [Medline: 10938172]
5. Wells PS, Anderson DR, Rodger M, Ginsberg JS, Kearon C, Gent M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: increasing the models utility with the SimpliRED D-dimer. Thromb Haemost 2000 Mar;83(3):416-420. [Medline: 10744147]
6. Franzen C. The Verge. 2013 Jun 3. Side effects may vary: the growing problem of unregulated medical apps URL: http://www.theverge.com/2013/6/3/4380244/how-should-medical-apps-be-regulated [accessed 2014-01-21] [WebCite Cache ID 6Mo0FAIF1]
7. Haffey F, Brady RR, Maxwell S. A comparison of the reliability of smartphone apps for opioid conversion. Drug Saf 2013 Feb;36(2):111-117. [doi: 10.1007/s40264-013-0015-0] [Medline: 23322549]
8. Morris R, Javed M, Bodger O, Gorse SH, Williams D. A comparison of two smartphone applications and the validation of smartphone applications as tools for fluid calculation for burns resuscitation. Burns 2013 Nov 15:1. [doi: 10.1016/j.burns.2013.10.015] [Medline: 24246618]
9. American College of Physicians. 2014. What is internal medicine? URL: http://www.acponline.org/about_acp/defining_internal_medicine/ [accessed 2014-01-21] [WebCite Cache ID 6Mo0bLMyP]
10. QxMD. URL: http://www.qxmd.com/ [accessed 2014-01-14] [WebCite Cache ID 6McZi4Ox3]
11. Medicon Applications. URL: http://mediconapps.com/ [accessed 2014-01-14] [WebCite Cache ID 6McaIwvLW]
12. Epocrates. URL: http://epocrates.com/ [accessed 2014-01-14] [WebCite Cache ID 6McaSLw7D]
13. MedCalc (Pro). URL: http://medcalc.medserver.be/ [accessed 2014-01-14] [WebCite Cache ID 6McaZIX8F]
14. medlive.cn. URL: http://www.medlive.cn/ [accessed 2014-01-14] [WebCite Cache ID 6McafnZ7Y]
15. Avivonet. URL: http://avivonet.com/ [accessed 2014-01-14] [WebCite Cache ID 6McakhQtB]
16. Irtza's Web Page. URL: http://irtza.com/ [accessed 2014-01-14] [WebCite Cache ID 6McaqgEq3]
17. ScyMed Medical Decision Systems. URL: http://scymed.com/ [accessed 2014-01-14] [WebCite Cache ID 6McawS4Fh]
18. Regular Rate & Rhythm. medimath URL: http://regularrateandrhythm.com/apps/medimath/ [accessed 2014-01-14] [WebCite Cache ID 6Mcb5UCyT]
19. Mediquations. URL: http://www.mediquations.com/iPhone_iPodTouch_iPad/index.html [accessed 2014-01-14] [WebCite Cache ID 6McbAkjup]
20. Medscape. URL: http://www.medscape.com/ [accessed 2014-01-14] [WebCite Cache ID 6McbFRSrR]
21. Vu C. iTunes. 2013. MedSolve Medical Calculator URL: https://itunes.apple.com/ca/app/medsolve-medical-calculator/id351057430?mt=8 [accessed 2014-01-14] [WebCite Cache ID 6McbSRv2r]
22. Skyscape. URL: http://www.skyscape.com/index/home.aspx [accessed 2014-01-14] [WebCite Cache ID 6McbwBzKv]
23. UpToDate. URL: http://www.uptodate.com/home [accessed 2014-01-14] [WebCite Cache ID 6Mcc0sZ24]
24. Manhattan Research. 2009. Physician smartphone adoption rate to reach 81% in 2012 URL: http://manhattanresearch.com/News-and-Events/Press-Releases/physician-smartphones-2012 [accessed 2014-01-21] [WebCite Cache ID 6Mo0kOkPF]
25. Glenn B. Medical Economics. 2013 Jun 13. Physicians' top 5 most-used medical apps for smartphones and tablets URL: http://medicaleconomics.modernmedicine.com/medical-economics/news/physicians-top-5-most-used-medical-apps-smartphones-and-tablets [accessed 2014-01-21] [WebCite Cache ID 6Mo0rUCHk]

26. Solomon S. Parkhurst Exchange. 2009 Nov. The 7 best medical smartphone apps URL: http://www.parkhurstexchange.com/practicemanagement/nov09/smartphone [accessed 2014-01-21] [WebCite Cache ID 6Mo10ZuVi]
27. Patel A. iMedicalApps. 2011 May 09. Top 10 iPhone medical apps for internal medicine physicians and residents URL: http://www.imedicalapps.com/2011/05/top-iphone-medical-apps-internal-medicine-physicians-residents/ [accessed 2014-01-21] [WebCite Cache ID 6Mo15O7lh]
28. US Food and Drug Administration. 2013 Feb 08. Is the product a medical device? URL: http://www.fda.gov/medicaldevices/deviceregulationandguidance/overview/classifyyourdevice/ucm051512.htm [accessed 2014-01-21] [WebCite Cache ID 6Mo1Ct4gb]

Abbreviations

CCS: Canadian Cardiovascular Society
CIWA: Clinical Institute Withdrawal Assessment
COPD: chronic obstructive pulmonary disease
FDA: Food and Drug Administration
MELD: Model for End-Stage Liver Disease
NSTEMI: non-ST elevation myocardial infarction
PE: pulmonary emboli
TIMI: thrombolysis in myocardial infarction
UNOS: United Network for Organ Sharing

Edited by G Eysenbach; submitted 31.10.13; peer-reviewed by K Payne, H Wharrad; comments to author 23.12.13; accepted 09.01.14; published 03.02.14.

Please cite as:
Bierbrier R, Lo V, Wu RC
Evaluation of the Accuracy of Smartphone Medical Calculation Apps
J Med Internet Res 2014;16(2):e32
URL: http://www.jmir.org/2014/2/e32/
doi:10.2196/jmir.3062
PMID:24491911

Original Paper

Google Glass for Documentation of Medical Findings: Evaluation in Forensic Medicine

Urs-Vito Albrecht[1*], MD, MPH; Ute von Jan[1*], PhD; Joachim Kuebler[2*], MD; Christoph Zoeller[2*], MD; Martin Lacher[2*], MD, PhD; Oliver J Muensterer[3*], MD, PhD; Max Ettinger[4*], MD; Michael Klintschar[5*], MD, PhD; Lars Hagemeier[5*], MD

[1]PL Reichertz Institute for Medical Informatics, Hannover Medical School, Hannover, Germany
[2]Department of Pediatric Surgery, Hannover Medical School, Hannover, Germany
[3]Division of Pediatric Surgery, New York Medical College, Valhalla, NY, United States
[4]Trauma Department, Hannover Medical School, Hannover, Germany
[5]Institute for Legal Medicine, Hannover Medical School, Hannover, Germany
[*]all authors contributed equally

Corresponding Author:
Urs-Vito Albrecht, MD, MPH
PL Reichertz Institute for Medical Informatics
Hannover Medical School
Carl-Neuberg-Str 1
Hannover, 30625
Germany
Phone: 49 5115323508
Fax: 49 5115322517
Email: albrecht.urs-vito@mh-hannover.de

Abstract

Background: Google Glass is a promising premarket device that includes an optical head-mounted display. Several proof of concept reports exist, but there is little scientific evidence regarding its use in a medical setting.

Objective: The objective of this study was to empirically determine the feasibility of deploying Glass in a forensics setting.

Methods: Glass was used in combination with a self-developed app that allowed for hands-free operation during autopsy and postmortem examinations of 4 decedents performed by 2 physicians. A digital single-lens reflex (DSLR) camera was used for image comparison. In addition, 6 forensic examiners (3 male, 3 female; age range 23-48 years, age mean 32.8 years, SD 9.6; mean work experience 6.2 years, SD 8.5) were asked to evaluate 159 images for image quality on a 5-point Likert scale, specifically color discrimination, brightness, sharpness, and their satisfaction with the acquired region of interest. Statistical evaluations were performed to determine how Glass compares with conventionally acquired digital images.

Results: All images received good (median 4) and very good ratings (median 5) for all 4 categories. Autopsy images taken by Glass (n=32) received significantly lower ratings than those acquired by DSLR camera (n=17) (region of interest: $z=–5.154$, $P<.001$; sharpness: $z=–7.898$, $P<.001$; color: $z=–4.407$, $P<.001$, brightness: $z=–3.187$, $P=.001$). For 110 images of postmortem examinations (Glass: n=54, DSLR camera: n=56), ratings for region of interest ($z=–8.390$, $P<.001$) and brightness ($z=–540$, $P=.007$) were significantly lower. For interrater reliability, intraclass correlation (ICC) values were good for autopsy (ICC=.723, 95% CI .667-.771, $P<.001$) and postmortem examination (ICC=.758, 95% CI .727-.787, $P<.001$). Postmortem examinations performed using Glass took 42.6 seconds longer than those done with the DSLR camera ($z=–2.100$, $P=.04$ using Wilcoxon signed rank test). The battery charge of Glass quickly decreased; an average 5.5% (SD 1.85) of its battery capacity was spent per postmortem examination (0.81% per minute or 0.79% per picture).

Conclusions: Glass was efficient for acquiring images for documentation in forensic medicine, but the image quality was inferior compared to a DSLR camera. Images taken with Glass received significantly lower ratings for all 4 categories in an autopsy setting and for region of interest and brightness in postmortem examination. The effort necessary for achieving the objectives was higher when using the device compared to the DSLR camera thus extending the postmortem examination duration. Its relative high power consumption and low battery capacity is also a disadvantage. At the current stage of development, Glass may be an adequate tool for education. For deployment in clinical care, issues such as hygiene, data protection, and privacy need to be addressed and are currently limiting chances for professional use.

(J Med Internet Res 2014;16(2):e53) doi:10.2196/jmir.3225

KEYWORDS

Google Glass; forensic medicine; autopsy, postmortem examination; documentation

Introduction

Background

Emerging technologies originally developed for the customer sector often find their way into professional environments. A prime example is the use of smartphones and tablet computers in combination with medicine-related apps in hospitals [1-3]. Google Glass [4] is another new and promising device originally developed with private consumers in mind that will soon be available for the general public. In principle, Glass is an optical head-mounted display that in addition to its technical capabilities also allows easy communication with various Internet-based services (mainly provided by Google). With some exceptions, its voice and gesture control functionalities allow an almost hands-free mode of operations. Using a prism, the display information is presented on the retina of the user and provides various kinds of information and visual feedback. The device is capable of taking pictures and recording videos by using an integrated camera. By using wireless access to the Internet, this information can be shared with the public. Also, because Glass provides a built-in microphone and a bone conduction transducer for audio signals, voice over Internet protocol (VoIP) communication is easily possible (eg, by using Google Hangout). Various sensors integrated into the device allow using it for rich augmented reality-based applications.

Currently, Glass is in a premarket state and only participants of the Glass Explorer Program have had the chance to evaluate the device so far [5], but some reports of proof of concept projects, especially in the medical field, are already available [6,7]. Lucien Engelen, based at Singularity University (Silicon Valley, CA, USA) and in Europe at Radboud University Medical Center, was the first health care professional in Europe to commence research on the usability and impact of Google Glass in the field of health care [8]. His collection of reports published since July 2013 comprises descriptions of experiences gained while using the device in a number of settings, including in operating theaters, ambulances, general practices, and other settings. His reports have been published primarily on social media channels [9].

At the moment, no empirical evidence for using Glass in a medical setting is known to the authors. A literature search on PubMed using the keywords “Google Glass” only resulted in a limited number of articles [10,11]. Nevertheless, Glass seems to be perfect for hands-free documentation purposes. Therefore, we were interested in the feasibility of integrating the device in a medical setting. We were also interested in usability aspects, including its effectiveness (can users successfully achieve their objectives?), efficiency (how much effort does it take to achieve those objectives?), and satisfaction (was the subjective experience satisfactory?). We decided to use Glass in a medical field where it was possible to evaluate the device with respect to these questions in a real setting while avoiding any risk to patients. The field of forensic medicine was selected as ideal for our purpose because photo documentation is of high value in this field.

Forensic Medicine

There are a number of demands placed on documentation in forensics. The description must be unambiguous and of high diagnostic value, yet easy to comprehend for laypersons, especially in court proceedings [12]. Photographic documentation may make a significant contribution to this and is expressly called for in the American Forensic Autopsy Performance Standards [13]. Because a textual description of a specific situation, such as anatomical structures, wounds, or hematomas, often requires imagination as well as a good knowledge of anatomy, a photograph may provide valuable help and aid readers in interpretation of the findings, especially if a reference scale is used. During forensic examinations and in autopsies, photographs can be used to document different stages of preparation if the original situation cannot be preserved. Particular findings can be recorded easily, in a timely manner, and are also easily verifiable and cost-efficient because of digital photography. However, when using conventional digital cameras, somebody has to operate the camera and this person must have a minimum of knowledge in camera handling and in the basics of forensic pathology. Until a few years ago, professional photographers were employed for this purpose, but nowadays—at least in Germany—this work has been delegated to the medical examiners themselves for economic reasons. The main problem with this approach is that an examiner who is working on a corpse cannot take hands-free photographs (Figure 1). Therefore, one has to change gloves to take pictures, often multiple times during the procedure, which costs time and resources. In this context, a camera that could be operated hands-free while taking the pictures and capturing the details the examiner wants with an image quality comparable to a commonly used digital single-lens reflex (DSLR) camera would mark an innovation in this field.

Figure 1. A typical crowded autopsy situation. On the left, the forensic pathologist is wearing Google Glass, allowing him to take pictures in a hands-free manner. On the right, the procedure is interrupted by an assistant taking a picture with a camera, which narrows the already limited space around the body.

Postmortem Examination and Autopsy

Postmortem examinations can be traced far back throughout history. Although the most important reason to perform such examinations is to determine whether a person is really dead—otherwise resuscitation measures must be initiated—the cause and manner of death (eg, natural death, accident, suicide, murder) has always been of further interest [13,14].

During a forensic examination of deceased persons, the examiner looks for externally visible evidence that may provide a hint to the manner of death, such as injuries, and samples may also be taken. In Germany, there is also a legal requirement to perform a forensic examination on every corpse that is to be cremated. This is because most indications and traces as to the cause of death will be permanently lost after cremation. Still, the results of a forensic examination have their limitations and the cause of death determined in this manner will always just be a suspected diagnosis. A forensic examination will never be able to substitute for an autopsy or have its probative value.

Over 100 years ago, Rudolf Virchow urged doctors to follow a standardized procedure for autopsy cases, which is meant as a useful guidance [15]. Nowadays, this is commonly acknowledged, but an autopsy should not be pressed into a rigid scheme [13,16,17]. Still, it should consist of 2 components, specifically an outer and an inner autopsy [18]. After describing the general impression, a systematic description of all body parts is required [13,19]. For this purpose, the head, throat and neck, chest, abdominal wall, back, outer genitalia and anus, as well as the upper and lower extremities are surveyed. Only when this is completed is the inner autopsy started. In Germany, §89 StPO (Code of Criminal Procedure) stipulates that all body cavities (head, chest, and abdominal cavity) must be opened and inspected, even if the cause of death seems to already have been ascertained during earlier steps. Every individual organ has to be examined and dissected [13,19]. If necessary, medical implants or foreign bodies as well as fractures, hematomas, and the skin, may have to be dissected to answer specific additional questions. In other jurisdictions (eg, following the American Forensic Autopsy Performance Standards), autopsies are performed in a similar manner [13].

Methods

Image Acquisition

We used Glass during one autopsy (Figures 1 and 2). The senior physician leading the autopsy was equipped with Glass and was instructed to perform the procedure following the previously described standards. Beforehand, the physician was trained in using the device and a self-developed app used for capturing images (which we called “Blink-app”), which took approximately 5 minutes in total. Because the device itself is very straightforward to use and the app only requires a single voice command to start and a nodding motion or a head shake (for keeping or deleting an image) once the image is taken, the physician stated he felt confident in using it after he had practiced taking images a few times.

In addition, 16 postmortem examinations of 4 cases were performed by 2 physicians with training in forensic medicine. Both physicians conducted their examinations alternately with Glass (8 examinations) and a DSLR camera (Olympus E-600, lens: Olympus *Zuiko* Digital ED 14-42 mm F3.5-5.6, 8 examinations) without using the internal flash of the camera. The kind of device and who began the examination on which case was randomized by running a random number generator (with numbers 0 and 1) 3 times: first, to determine who of the 2 examiners should start, the second time to determine the device to use for the first of the 2 examinations of that examiner. The third call was used to determine the order of devices for the second examiner. The forensic pathologists were asked to follow their standardized routine. Therefore, they were not required to take a specific number of pictures, but were allowed to take as many pictures as they deemed necessary for documentation purposes.

Figure 2. A forensic pathologist taking a picture of the situs using Google Glass during an autopsy (left). The Blink-app app transmits the image to the laptop where it is immediately presented to the attending detective and the public prosecutor (right).

Glass

The device—a Google Glass explorer version—available during our prestudy (performed at the beginning of December 2013) ran on Android 4.0.4 (XE11). At the time of the prestudy, the Glass development kit had become available to the general public, which enabled us to build applications using methods and tools already established for other Android-based devices. Specifications of the available developer explorer unit included a Texas Instruments open multimedia applications platform (OMAP) 4430 SoC 1.2G Hz Dual (ARMv7), a 5-megapixel camera, 682 MB of memory, and 16 GB of storage, of which 12 GB were available for user purposes, as well as Wi-Fi 802.11b/g and Bluetooth. Sensors, such as a gyroscope, an accelerometer, and a magnetometer (all for 3 axes), as well as an ambient light and proximity sensors were also integrated in the device [20].

Software Setup

For image acquisition, we decided to refrain from using the preinstalled camera app for two reasons. First of all, without having performed an analysis of the data streams that occur whenever the device logs into a wireless network, we were unsure whether data acquired using the stock camera app would be relayed in any way to any third parties, such as Google. This would be understandable from their point of view given that the device available for our study was still at a premarket stage and they would want to obtain data about possible bugs. Nevertheless, although we were dealing with deceased persons, this would still have been problematic because of the sensitive nature, both ethically and legally, of the procedures that were performed. Secondly, at the time of our prestudy, running the preinstalled app completely hands free was problematic. Nevertheless, because of hygiene issues, this is a necessity in most medical environments, including forensic postmortem examinations and autopsies.

Therefore, using the available development kit, a native app was developed that alleviated these problems. This app, called Blink-app, made use of the standard camera application programming interface (API) provided by Google. It could be started through voice commands and used specific but easy to perform gestures, such as nodding or shaking the head (evaluated using rotation vectors acquired from the geomagnetic sensor), to allow the user to specify whether an image should be kept or deleted. Using functionality integrated into this app, all accepted images were stored in a separate folder on the device and were securely transmitted via Wi-Fi to a Linux-based laptop that served as a wireless hotspot for Glass, but did not have any connection to the Internet. On this laptop, an Apache server had been set up to accept and store images transmitted from the device through hypertext transfer protocol (HTTP) POST requests for later evaluation. Additionally, the server setup allowed other personnel to review the acquired images in a Web browser on the laptop’s (larger) screen immediately after they were taken. Image viewing could be performed in various scale modes ranging from an overview to zooming into the image (up to its original resolution) to allow onlookers to closely scrutinize specific areas. As soon as the procedures were over, any images remaining on the device were manually deleted.

Usability and Acceptance

The 2 forensic pathologists involved in taking the images were interviewed about usability aspects of Glass in combination with the Blink-app app. We asked questions about general experience, usage, and handling of the device, noteworthy positive and negative aspects, and whether there were any surprises. Because only 2 examiners used Glass, we refrained from using standardized usability instruments.

Subjective Ratings of Image Quality

A total of 6 forensic pathologists (3 male, 3 female; age range 23-48 years, mean age 32.8 years, SD 9.6), with mean work experience of 6.2 years (SD 8.5) were included in this evaluation after having given their informed consent for participation in the evaluation of the quality of the acquired images. The evaluation was done in two parts: the images were divided into 2 groups according to the settings they were taken in (autopsy and postmortem examination) and for each of the images, the device used for acquisition was noted (DSLR camera or Glass) in the internal database of the app we used for presenting and evaluating the images. All pictures were evaluated by all 6 participating forensic pathologists. To avoid differences in presentation, all used the same tablet computer with fixed display settings (maximum brightness and automatic adjustment for brightness had been deactivated) under similar lighting conditions. Also, during their evaluation, the raters were not informed about which of the two devices had been used for taking the presented images.

In step 1, 49 pictures of the autopsy were presented to the participants using a self-developed evaluation app. The images taken by both devices were intermixed and were then presented in a randomized manner on a single 10” Android-based tablet computer (Samsung Galaxy Note 10.1N).

The participants were presented each image sequentially and they had to give their opinion about whether they were satisfied with it with respect to 4 parameters:

1. Region of interest or; specifically, whether all necessary anatomical structures were depicted.
2. Sharpness.
3. Overall color setting; specifically, whether the images allowed adequate discrimination of even small changes in color which could be important to document hints about underlying pathologies (eg, bruises on a decedent’s skin). We were not interested in color cast caused by the devices.
4. Image brightness achievable based on the defined lighting installed in the autopsy room. Although one could argue that using a flash might have improved brightness of the images taken with the DSLR camera (and give an advantage to the DSLR camera because Glass does not include a flash), a flash is rarely employed during autopsies or forensic examinations because it could easily cause overexposure at close distances and could also lead to reflections when photographing wet tissue during an autopsy.

Raters were asked to specify their opinion about each image using a Likert scale with 5 levels (ie, ++, +, +/-, -, and --, in which “++” represented a highly positive and “--” represented an extremely negative rating for the respective parameter. For statistical analysis, these values were transformed into numeric values, in which the highest possible rating “++” corresponded to 5 and the lowest rating “--” corresponded to 1.

In step 2, using the images acquired during the postmortem examinations, we were primarily interested in whether there were significant differences in how well users captured the desired anatomy when using either the DSLR camera or Glass. Because one device is handheld and the other is head mounted, the way users aim the device and shoot the image differs. Because there were small but perceivable differences in coloring (specifically, a very slight yellowish tint in the DSLR camera images), we decided to apply an automatic white balancing algorithm integrated into the GNU Image Manipulation Program (GIMP) 2.8 for all images of the postmortem examination. Based on the RGB color model, for each color channel, this algorithm discards pixel values at both ends of the histogram for the respective channel that contributes only to 0.05% of the image and stretches the remaining pixel values as much as possible. This procedure avoids undue influence of outliers at both ends of the spectrum for each of the channels [21]. For documentation purposes in forensics, this is a commonly applied method; thus, it does not add any steps. But, in our case, the algorithm was specifically applied because we did not want the raters’ decisions about which device was used for each of the presented images to be biased by the differences in tint. With respect to color, as stated previously, we were interested in whether the devices allowed adequate discrimination of colors. Independent of the device used for image acquisition, the algorithm was used on all images acquired during the postmortem examinations without informing the participants about this process before they gave their ratings. After performing the white balance for all 110 pictures of the postmortem examination, the images were loaded into the aforementioned app as described previously and presented to each of the participants in a randomized manner.

Statistical Analysis

An observer recorded the time span (in seconds) required for each single postmortem examination using the stopwatch functionality available on a separate smartphone (Samsung Galaxy S4). The physicians had to start their examinations from a defined location on the bodies to be examined. A Wilcoxon signed rank test [22] was used to compare the related samples to detect differences between the devices. For each of the postmortem examinations, the number of images taken with the device chosen for that examination (Glass or DSLR camera) was counted. Before and after each use of Glass during an examination, the remaining battery charge was noted. For all mentioned variables, descriptive statistics were calculated, including the mean and standard deviation (SD).

The descriptive statistics for the ratings included the tabulation of the frequency and percentages of scale items for each item per device. Median values and quartiles were also calculated. To detect differences between the ratings obtained for both devices, we calculated an unpaired rank sum test (2-sided Mann-Whitney *U* test, with Cronbach alpha=.05 [23]). All items were included and there were no missing entries.

To determine interrater reliability (ie, how strongly the ratings of the participants correlated), we calculated intraclass correlation coefficients (ICC) [24] for the 6 participants for the items region of interest, sharpness, color, and brightness. The ICC can be used to assess the consistency of quantitative measurements (ie, correlation) between multiple observers measuring the same quantity [25]. All observers rated each case—they were not randomly chosen. We considered single values of the observers. We decided to use the ICC (3,1)-type, 2-way mixed with average measures for the calculations that were conducted using SPSS 21 (IBM Corp, Armonk, NY, USA).

Institutional Review Board Approval

The study was conducted with approval by the Institutional Review Board of Hannover Medical School, study number 2069-2013.

Results

Image Acquisition

During the autopsy, 64 pictures were taken: 40 with Glass and 24 using the DSLR camera. A total of 15 images were excluded because they were not related to the deceased person’s anatomy, but rather to pictures of additional paperwork provided by the authorities. Thus, 49 autopsy images (Glass: n=32; DSLR camera: n=17) were used for the study. During the postmortem examinations, 112 pictures were taken, 55 of these using Glass and 57 using the DSLR camera. We excluded 2 pictures (for similar reasons); thus, 110 images (Glass: n=54; DSLR camera: n=56) remained for evaluation.

During postmortem examinations, an average number of 7 pictures were taken per case with both devices. When using the DSLR camera, the mean duration of a single postmortem examination was 225.9 seconds (SD 50.6) compared to Glass, for which the mean duration was 268.5 seconds (SD 64.1) (Table 1). During the postmortem examinations, an average 5.5% (SD 1.85) of Glass' battery charge was used per case, corresponding to 0.81% per minute or 0.79% per picture.

Table 1. Descriptive statistics of time measurements, number of pictures taken (per case), and loss of battery charge (per case) during postmortem examinations, stratified by device used to capture image: digital single-lens reflex (DSLR) camera or Google Glass.

Usage statistics per device	Minimum	Maximum	Mean (SD)
Time measurements (seconds)			
DSLR camera	189	310	225.9 (50.6)
Glass	180	359	268.5 (64.1)
Pictures taken			
DSLR camera	5	11	7.25 (2.12)
Glass	4	9	7.0 (1.85)
Used percentage of battery capacity per case (%)			
DSLR camera	N/A[b]	N/A	N/A
Glass	3	8	5.5[a](1.85)

[a] Battery decrease per minute: 5.5%/(268.5 s/60 s)=0.81%; per picture: 5.5%/7.0 pictures=0.79%.

[b] N/A: not applicable

Interviews

Based on the interviews, we obtained subjective ratings of the user experience for both Glass and the Blink-app app. In the forensic setting, especially during the autopsy, Glass equipped with Blink-app was deemed as a suitable tool for examiners in situations where they needed both of their hands for fulfilling tasks, especially in cases where there is limited opportunity for other persons to take pictures (either because of space requirements or availability). This certainly holds true in autopsy settings. Another useful effect was that other persons attending the procedure (colleagues and a police officer) were directly able to review what the leading pathologist had seen by looking at the laptop's screen (Figure 2). In the version that was tested, for users who do not have to wear corrective glasses, the device does not disturb the examiner's sight during the procedure and is comfortable to wear because of its low weight and good ergonomics. Even in a busy environment, such as an autopsy with 5 people involved, the voice command used for starting the app worked very well. The integrated gesture control (nodding for upload and shaking one's head for discarding an image) was perceived as natural.

Similarly, for the postmortem examinations, both examiners agreed that the device's ergonomics and light weight made it comfortable to wear. Because the environment is generally quieter when performing this kind of procedure, both the voice and gesture control worked well. The examiners mentioned that taking images using Glass took more physical effort than with the DSLR camera to capture the desired regions of interest (Figure 3). This was especially the case with close-ups because the device is placed on the head during use and there was no macro function available. Consequently, to obtain images of the region they wanted, they sometimes had to bring their head—and with it the device—closer to the findings than they would have preferred.

Figure 3. Similar phases of a postmortem on the same body with photo documentation (image anonymized for patient confidentiality). The image on the left was taken with Google Glass and the image on the right with the DSLR camera.

Rating of the Images

The 6 raters evaluated 49 autopsy images each, totaling 294 evaluations (DSLR camera: n=102; Glass: n=192), each consisting of ratings on a 5-point Likert scale (from "++" to "--" corresponding to a numeric scale of 5 to 1) for each of the 4 qualities region of interest, sharpness, color, and brightness. Table 2 shows the absolute frequencies and percentages obtained per quality for each of the scale items for both devices. Differences between both devices are particularly noticeable for region of interest and sharpness. For region of interest, a large percentage of images taken by DSLR camera achieved the highest rating of "++" (region of interest: 57.8%, sharpness: 68.6%), whereas for images taken by Glass, ratings for these 2 qualities were more evenly distributed with a somewhat smaller peak for rating "+" (region of interest: 41.7%, sharpness: 34.9%). The distributions of ratings for color and brightness for both devices were a closer match, although results were also slightly in favor of images taken by DSLR camera.

For the images taken during the postmortem examinations (n=110), a total of 660 single evaluations were obtained from the 6 raters (DSLR camera: n=342; Glass: n=318). Each of the 4 qualities (region of interest, sharpness, color, and brightness) were rated using a 5-point Likert scale (from "++" to "--" corresponding to a numeric scale from 5 to 1). Table 2 shows the absolute frequencies and percentages obtained per quality for each of the scale items for both devices. Considering the percentage values for the scale items for sharpness, color, and brightness, ratings are only slightly in favor of the DSLR camera. For region of interest, the difference is considerable, as shown when looking at the sum of ratings for "+" and "++" (corresponding to positive and highly positive ratings), which amounted to 90.7% for images acquired using the DSLR camera vs 61% for those taken by Glass.

As mirrored by the median values and interquartile ranges (IQR) shown in Table 3, for both devices, the 6 raters were quite consistent in their evaluation of the 4 qualities (region of interest, sharpness, color, and brightness) for images taken during the autopsy (n=49). For the postmortem examinations, raters were again quite consistent in their evaluation of the 4 qualities (region of interest, sharpness, color, and brightness) (Table 3).

Table 2. Frequencies (n) and percentages (%) of evaluations given by 6 raters for images taken during autopsy (n=294) and postmortem examination (n=660) with a digital single-lens reflex (DSLR) camera and Google Glass.

Quality and rating scale[a]	Autopsy evaluations, n (%)		Postmortem examination evaluations, n (%)	
	DSLR camera (n=102)	Glass (n=192)	DSLR camera (n=342)	Glass (n=318)
Region of interest				
++	59 (57.8)	56 (29.2)	161 (47.1)	82 (25.8)
+	33 (32.4)	80 (41.7)	149 (43.6)	112 (35.2)
+/-	7 (6.9)	35 (18.2)	28 (8.2)	92 (28.9)
-	3 (2.9)	20 (10.4)	4 (1.2)	32 (10.1)
--	0 (0)	1 (0.5)	0 (0)	0 (0)
Total	102 (100)	192 (100)	342 (100)	318 (100)
Sharpness				
++	70 (68.6)	45 (23.4)	160 (46.8)	146 (45.9)
+	24 (23.5)	67 (34.9)	129 (37.7)	111 (34.9)
+/-	6 (5.9)	37 (19.3)	37 (10.8)	46 (14.5)
-	2 (2.0)	40 (20.8)	9 (2.6)	15 (4.7)
--	0 (0)	3 (1.6)	7 (2.0)	0 (0)
Total	102 (100)	192 (100)	342 (100)	318 (100)
Color				
++	38 (37.3)	40 (20.8)	134 (39.2)	147 (46.2)
+	52 (51.0)	86 (44.8)	152 (44.4)	88 (27.7)
+/-	11 (10.8)	54 (28.1)	47 (13.7)	71 (22.3)
-	1 (1.0)	12 (6.3)	9 (2.6)	12 (3.8)
--	0 (0)	0 (0)	0 (0)	0 (0)
Total	102 (100)	192 (100)	342 (100)	318 (100)
Brightness				
++	29 (28.4)	32 (16.7)	124 (36.3)	129 (40.6)
+	47 (46.1)	79 (41.1)	147 (43.0)	102 (32.1)
+/-	21 (20.6)	61 (31.8)	61 (17.8)	52 (16.4)
-	5 (4.9)	20 (10.4)	10 (2.9)	34 (10.7)
--	0 (0)	0 (0)	0 (0)	1 (0.3)
Total	102 (100)	192 (100)	342 (100)	318 (100)

[a] Rating “++” indicates a highly positive rating; “--”stands for very poor results.

Table 3. Median values and interquartile ranges (IQR) of evaluations given by 6 raters for images taken during autopsy (n=294) and postmortem examination (n=660) with a digital single-lens reflex (DSLR) camera and Google Glass.

Quality and participant	Autopsy evaluations[a]				Postmortem examination evaluations[a]			
	DSLR camera (n=102)		Glass (n=192)		DSLR camera (n=342)		Glass (n=318)	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR
Region of interest								
#01	5	1	4	1	5	1	5	1
#02	5	1	4	1	5	1	4	1
#03	5	0	5	0	5	0	5	1
#04	4	0	4	1	4	0	3	1
#05	4	2	3	2	4	0	3	1
#06	4	1	3	2	4	2	3	1
Sharpness								
#01	5	0	4	1.25	4	1	5	1
#02	5	0	4	1	4	1	4	1
#03	5	0	4	1	5	0	5	0
#04	4	1	2.50	2	4	0	4	1
#05	5	0	4	2	5	1	5	0
#06	5	1	3	2	4	2	4	1
Color								
#01	4	1	4	0	4	1	5	1
#02	4	1	3	0.50	3	1	3	0
#03	5	1	4	1	5	0	5	0
#04	4	0	3	1	4	0	4	1
#05	5	1	4	1	5	0	5	0
#06	5	1	4	1	4	1	4	1
Brightness								
#01	4	0	4	0	4	0	4	1
#02	3	1	3	0	3	1	3	2
#03	5	0	5	1	5	0	5	0
#04	3	1	3	1.25	4	1	3	2
#05	4	1	4	1	5	0	5	0
#06	4	1	3	1	4	1	4	1

[a] For calculating median and IQR, the Likert scale items have been transformed into a numeric representation between 1 and 5 where 5 represents the best (“++”) and 1 the worst possible rating (“--”).

For image quality, it was of interest whether the region of interest was appropriately captured, whether the image was well focused and sharp, as well as whether color discrimination and brightness were satisfactory for the participants. Instead of using algorithms for an objective interpretation of the images, we decided to obtain subjective ratings by professionals who would have to use such images for their photo documentation if a device such as Glass were to be officially introduced in such settings. Overall, on a 5-point scale (where 5 represented the best and 1 the worst possible rating) the images received good (median 4) and very good values (median 5) in all 4 categories independent of the device used for image acquisition or the setting (Table 4).

Table 4. Minimum (min) and maximum (max) measurements, first quartile (Q1), second quartile (Q2, median), third quartile (Q3), and interquartile range (IQR) of ratings from 6 raters for images taken during autopsy (images: n=49, ratings: n=294) and postmortem examination (images: n=110, ratings: n=660) with a digital single-lens reflex (DSLR) camera and Google Glass.

Procedure and measurements	Region of interest		Sharpness		Color		Brightness	
	DSLR camera	Glass	DSLR camera	Glass	DSLR camera	Glass	DSLR camera	Glass
Autopsy								
Min	2	1	2	1	2	2	2	2
Max	5	5	5	5	5	5	5	5
Q_1	4	3	4	3	4	3	3	3
Q_2[a]	5	4	5	4	4	4	4	4
Q_3	5	5	5	4	5	4	5	4
IQR[a]	1	2	1	1	1	1	2	1
Postmortem examination								
Min	2	2	1	2	2	2	2	1
Max	5	5	5	5	5	5	5	5
Q_1	4	3	4	4	4	3	4	3
Q_2[a]	4	4	4	4	4	4	4	4
Q_3	5	5	5	5	5	5	5	5
IQR[a]	1	2	1	1	1	2	1	2

[a] For calculating median and IQR, the Likert scale items have been transformed into a numeric representation between 1 and 5 where 5 represents the best (“++”) and 1 the worst possible rating (“--”).

Pictures taken during the autopsy showed differences between DSLR camera and Glass primarily in region of interest and sharpness (DSLR camera: median 5; Glass: median 4). In color and brightness, the median values for both devices were identical (median 4), but values for the third quartile differed because only the DSLR camera received maximum ratings of 5. Nevertheless, images taken during the autopsy using Glass received significantly lower ratings for all 4 categories than those taken by the DSLR camera (region of interest: *z*=–5.154, *P*<.001; sharpness: *z*=–7.898, *P*<.001; color: *z*= –4.407, *P*<.001, brightness: *z*=–3.187, *P*=.001; see Table 5). Raters favored the pictures taken by DSLR camera with respect to correctly capturing the desired region of interest as well as sharpness, color discrimination, and brightness.

Table 5. Unpaired rank sum, 2-sided Mann-Whitney *U* (Cronbach alpha=.05) for ratings of autopsy and postmortem examination images taken by digital single-lens reflex (DSLR) camera and Google Glass.

Item	Autopsy (n=294)		Postmortem examinations (n=660)	
	z	*P*	*z*	*P*
Region of interest	–5.153691	<.001	–8390	<.001
Sharpness	–7.898378	<.001	–0.587	.56
Color	–4.406570	<.001	–0.011	.59
Brightness	–3.186663	.001	–540	.01

Interrater Reliability

The interrater reliability was high (Table 6). The ICC for the ratings obtained based on the autopsy pictures indicated a strong positive relationship for sharpness (ICC=.917, 95% CI .875-.948, *P*<.001) and brightness (ICC=.720, 95% CI .579-.826, *P*<.001), and a moderately positive relationship between raters for color (ICC=.658, 95% CI .485-.787, *P*<.001) and region of interest (ICC=.630, 95% CI .443-.770, *P*<.001). When considering the ratings of the postmortem examinations, the ICC values for region of interest (ICC=.727, 95% CI .639-.799, *P*<.001) and sharpness (ICC=.761, 95% CI .685-.824, *P*<.001) indicate a strong positive relationship, whereas color (ICC=.674, 95% CI .569-.760, *P*<.001) and brightness (ICC=.545, 95% CI .399-.665, *P*<.001) indicate a moderately positive relationship among the raters.

Table 6. The intraclass correlation coefficients (ICC), 95% confidence intervals (95% CI), and levels of significance (*P*) for items rated by 6 raters for autopsy and postmortem examinations, ICC(3,1), 2-way mixed, average measure.

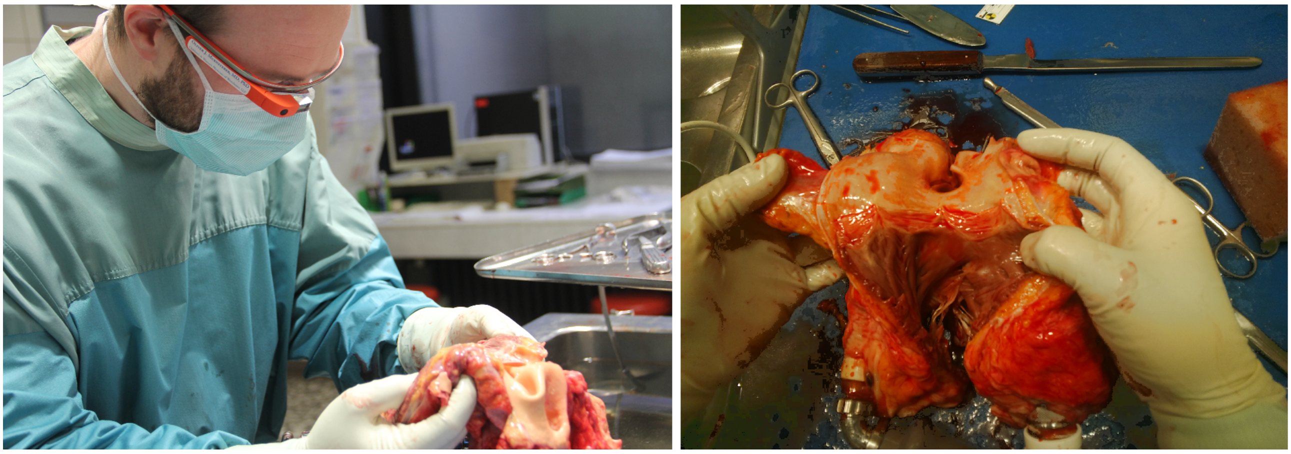
Item	Autopsy			Postmortem examinations		
	ICC	95% CI	*P*	ICC	95% CI	*P*
Region of interest	.630	.443-.770	<.001	.727	.639-.799	<.001
Sharpness	.917	.875-.948	<.001	.761	.685-.824	<.001
Color	.658	.485-.787	<.001	.674	.569-.760	<.001
Brightness	.720	.579-.826	<.001	.545	.399-.665	<.001

Discussion

Principal Findings

In our evaluation, we focused on the main functionality of Glass in areas where we thought it would have an advantage against other existing technology. Because it includes a lightweight head-mounted display and a camera with voice and gesture control in combination with embedded computer and wireless communication technology, it seems ideal for both communication and photo documentation (Figure 4). We were interested in its effectiveness, efficiency, and user satisfaction.

Figure 4. The forensic pathologist is taking a picture of the heart using Google Glass (left). The original image taken with the device allows viewers a first-person perspective of the structure (right).



Overall, both devices had sufficient capabilities and could be used effectively for the task of image acquisition during the procedures. Users were able to successfully achieve their objectives with either device, but we detected differences in the efficiency of the devices. Postmortem examinations performed using Glass for acquiring the photo documentation took 42.6 seconds longer than those performed with the DSLR camera (Table 1). The difference was significant. (z= –2.100, P=.04 using Wilcoxon signed rank test). Overall, this corresponds to one-fifth of the average length of the procedure when using a DSLR camera (3.45 min) and leads to additional undesired expenditure of battery power. On average, 5.5% (SD 1.85) of the battery capacity of Glass was used per case during the postmortem examinations, corresponding to 0.81% per minute or 0.79% per picture.

Although one may suspect that this relatively high power consumption may be because of not using the stock camera app for capturing the images, we do not believe this to be the case. While developing the app, we did not note any differences in power consumption when comparing our app to the preinstalled camera app. As noted previously, the Blink-app app was implemented based on the official camera API calls; thus, both apps made use of the same (or at least very similar) calls to image acquisition functions. Also, when using the Blink-app app, images were saved in the original form available through the official API calls without any alteration. Of course, power consumption is higher with an active Wi-Fi connection, but again, we noted no apparent difference between the stock camera app and the Blink-app app if the device was connected to a wireless hotspot while either app was running. Also, because we were dealing with a premarket device, we expect power consumption and battery capacity to improve in future versions. At the time of our study, our aim was simply to determine whether, in its current state, power consumption was a potential limiting factor or was adequate for our chosen setting without having to resort to external power sources, such as additional battery packs, because these would be problematic in such settings due to external cabling and additional bulk. In medical settings, a cable running down from Glass to an external battery pack may raise concerns with respect to hygiene as well as add potential for Glass to be pulled off the user’s nose if something (eg, a fastener on the physician’s surgical gown) inadvertently pulled on this cable.

For the DSLR camera, it was not possible to determine the exact percentage of battery power used because the camera only provided an icon with a crude scale in units of 25%; however, no significant loss of battery power could be detected. This is understandable when considering the average battery capacity of current DSLR cameras; thus, no corresponding information is listed in Table 1. For Glass, the battery's level of charge quickly decreased; an average 5.5% of the battery's capacity was spent per postmortem examination (0.8% per minute or 1.3% per picture). If all went well and examiners were to use Glass in a similar manner to how it was used in our study, this would allow for approximately 18 postmortem examinations or 125 minutes of use, or 78 pictures. Compared to the DSLR camera, this would require earlier recharging, which could be a problem if there were a large number of postmortem examinations to be performed. But it is expected that a future version of Glass sold to the general public will have an improved battery life. Altogether, the effort necessary for achieving the objectives was higher when using Glass compared to the DSLR camera, and when comparing the percentage of battery charge used, the numbers currently seem to be also in favor of the DSLR camera.

The presentation of the region of interest is the most important aspect that must be covered by the photographer and the device used. For the images taken during the forensic examinations, we decided to apply an automatic white balancing algorithm as described previously to give the images taken by both devices a similar and comparable appearance and to allow the participants to more easily focus on possible differences in the presentation of the region of interest. As expected, the median of the ratings for images taken with Glass or DSLR camera showed lower variability in the median (median 4) in all item categories because of the image manipulation. The IQR showed greater spans in the Glass group (IQR 2) with respect to region of interest, color, and brightness. Based on statistical testing, only the ratings for region of interest (z=–8.3901, P<.001) and brightness (z=–540, P=.01) were significantly better for images taken with the DSLR camera, whereas sharpness and color showed no significant differences (Table 5). The main disadvantage concerning region of interest is the lack of a zoom function with Glass compared to the DSLR camera.

User satisfaction depends on a number of factors, including the usability of the device and the quality of the images. The interviews about user experience and acceptance of the Glass device underline the comfort of a lightweight voice- and gesture-controlled device with a head-mounted camera. Those questioned perceived the device to be a suitable tool in the situations where they had used it (ie, during autopsy and forensic examinations).

Overall, the experience using Glass was satisfactory although the quality of the images obtained left room for improvement. Our use of the custom app instead of the device's stock camera did not appear to have an influence on image quality. While developing the app, a careful visual comparison of images taken with both apps (taken by the same person with same angle of view and similar lighting conditions because the images were taken at the same location and only a few seconds apart) did not show any obvious differences in color, sharpness, and other parameters. There were also no visual differences with respect to artifacts or depiction of fine structures. Additionally, an analysis of the exchangeable image file format (Exif) data included within the images did not show any notable differences (aside from the timestamp). Therefore, we believe that the quality of images acquired by Glass was not negatively influenced by our app. However, it is to be expected that the manufacturer will refine many of the points we noted (ie, image quality and battery capacity) before the device hits the consumer market.

Limitations

There are a number of limitations in our evaluation. First of all, because of time constraints, we were only able to use the device on a limited number of cases and with only 2 physicians. Also, for our study, the 6 raters included the 2 physicians who took the images, but because of the amount of images taken from similar perspectives with both devices by both physicians and some days having passed between image acquisition and evaluation, we do not believe this influenced the results. When asked, neither of them was able to identify who had taken a specific image because of the random manner in which the images were presented. In most cases, they were not even sure which case an image belonged to. Also, neither of the physicians had any influence on the integration of the selected images in the app we used for image presentation.

Because it was only possible to use Glass during one autopsy, we may have missed differences one might otherwise note with respect to specific types of cases that require other approaches than those commonly used in standard autopsy situations. Establishing a control group could also have improved the results. A larger number of postmortem examinations, ideally performed by additional examiners, would have reduced bias and standardized instruments for measuring usability could have been applied as well (eg, the system usability scale by Brooke [26] and Hassenzahl's AttrakDiff2 [27]). Raters recruited from other forensic medical facilities would also have improved the data pool used for analysis.

From a technical point of view, it may be seen as problematic to compare the capabilities of a DSLR camera with a resolution of 12.6 megapixels with those of a 5-megapixel camera integrated in a mobile device. Even with recent advances in mobile technology, the quality of a small-lensed camera can never compare to what a DSLR camera has to offer. However, our aim was not a direct comparison of technical parameters but to determine whether the perceived image quality provided by Glass was adequate for the stated purpose; therefore, we do not believe this to be a limitation.

Additional Advice

There are additional aspects that will make the deployment of Glass in clinical settings complicated. These expected complications are not due to purely technical issues, such as image quality or handling, but rather concerns about data protection and privacy. Because Glass was developed primarily for the private sector, its basic functionality and the preinstalled apps make extensive use of Google's network and servers. Therefore, the user has little control over the way data are handled, transmitted, stored, and possibly evaluated by a third party. Because Google is the main company involved, their data protection policy is applied. For the professional medical sector, it is not advisable to send any data of a patient—especially concerning medical issues—using an open and unsecured network. At least in European countries, it is not acceptable to store and share medical data stored in "the cloud" and similar restrictions apply to other countries (eg, the US Health Insurance Portability and Accountability Act does not allow a third party to access patient data [10]). Therefore, unencrypted communication or communication over insecure networks for making video calls or doing live chats to exchange information about a patient's case are just as unacceptable as sending pictures and emails containing personal details that help identify the patient.

Currently, when using Glass in a medical setting, a private (closed) network without any connection to servers aside from those belonging to the private infrastructure should be ensured. Additionally, just as for all other applications where medical data are concerned, state-of-the-art encryption and access policies should be employed for access to the infrastructure as well as storage and transmission of all data. For our study, we developed the previously mentioned app that allowed data exchange only within a private network (ie, between Glass and a laptop without Internet access).

There is also an additional point that must be kept in mind regarding apps running on Glass. Just as for other mobile smart devices used in health care settings, there will be additional pitfalls regarding data security and privacy once the market of third-party health apps and medical apps specifically adapted to Glass grows; for example, if manufacturers of apps do not implement appropriate measures for ensuring these aspects in their products. Depending on the jurisdiction they are used in, applications that have a diagnostic or therapeutic purpose (ie, could be rated as a medical device) must already conform to regulations (eg, the Mobile Medical Applications Guidance for Industry and Food and Drug Administration Staff [28]), thus ensuring some quality control. Nevertheless, for apps for which such regulations do not apply, it would be highly desirable if manufacturers or developers of an app were to provide users with at least sufficient and transparent information to allow them to make an informed decision about whether they want to use an app—be it on Glass or on other mobile devices—or not. The information should be provided in a clearly structured way (eg, using an app synopsis as presented by Albrecht [29]). This could also be used as a starting point for instigating a peer-review process of such applications.

Devices such as Glass can be used unobtrusively without attracting attention of persons who are unaware of the technical possibilities or dangers offered by these devices. As in all areas of medicine, the usual rules regarding a patient's rights of privacy need to be applied, which may easily be overlooked because Glass is so easy to use and integrates well in many settings. Also, because it is currently unknown how living patients will react to the device in an examination and other common settings encountered in the medical field, empirical studies that investigate the acceptance of patients would add significant value if Glass is to be widely deployed. For example, a recent blog entry by Mat Honan [30] who used the device for a year indicated issues with acceptance from the public.

Hygiene is also an issue when using the device with patients because Glass is not a medical product and was not developed for use in a sterile environment. Therefore, it is questionable whether it is possible to disinfect it properly. Currently, the manufacturer is not giving any recommendations regarding proper disinfection of the device. By allowing hands-free operation, our Blink-app app supports the demands of hygiene; nevertheless, we used disinfecting wipes for plastic surfaces of medical devices, although this may cause issues with warranty [31].

Conclusions

In our opinion, aside from using the device for documentation purposes, there is also potential for another field of application in forensic medicine: By chance, we recognized the reaction of our medical students participating in the autopsy. They were amazed by the pictures taken by Glass that were almost instantly presented on the laptop's screen. Because there is only room for a limited number of students to watch, even for interesting cases, using the described setup with our additional add-ons might provide an opportunity for a larger number of students to observe the procedure in a dynamic manner. There are also a number of other uses in medical education and training one can imagine (eg, cardiology and others described in [32]).

The main strength of Glass is its ergonomic and lightweight design, combined with the camera that allows taking pictures and videos directly from the user's point of view. This feature makes the design interesting for medical education (eg, in pathology and in surgery). Independent of the local setting, students and colleagues may join a complicated autopsy or operation and can see exactly what their teachers see, thus learning from watching their actions from the ideal perspective (Figure 4). This is similar to the potential benefit also already noted—albeit for education in general—by other authors [32-34]. If worn by students, Glass could also serve to enhance existing augmented reality-based solutions for medical education, such as mARble [35], an augmented reality-based blended learning tool for use in medical education that is currently implemented for conventional smartphones and allows students to immerse themselves in an almost realistic learning scenario in cases where learning on real patients may be restricted (eg, because of ethical concerns). It remains to be seen whether the current comparatively low display resolution of Glass is sufficient to provide students with content similar to what can currently be shown using the smartphone-based solution (eg, by overlaying specific medical findings on another student's skin). If yes, Glass would also have potential to significantly enhance augmented reality-based learning tools.

Acknowledgments

We would like to thank Reinhold Krentscher and Michael Erdmann (assistants to the forensic pathologists) as well as Roman Wolff-Maras for their support conducting this study. We thank Nuray Tasdelen, Christoph Noll, and Enno Rubner for assistance during the photo shooting. The study was financed solely from institutional funds of the Hannover Medical School. We acknowledge support by Deutsche Forschungsgemeinschaft for the publication costs.

Conflicts of Interest

None declared.

References

1. Migliore M. Smartphones or tablets for a better communication and education between residents and consultant in a teaching hospital. J Surg Educ 2013 Aug;70(4):437-438. [doi: 10.1016/j.jsurg.2013.03.001] [Medline: 23725929]
2. Székely A, Talanow R, Bágyi P. Smartphones, tablets and mobile applications for radiology. Eur J Radiol 2013 May;82(5):829-836. [doi: 10.1016/j.ejrad.2012.11.034] [Medline: 23312700]
3. Dexheimer JW, Borycki EM. Use of mobile devices in the emergency department. Stud Health Technol Inform 2013;192:1086. [Medline: 23920860]
4. Google Glass. 2014. URL: http://www.google.com/glass/start/ [accessed 2014-01-03] [WebCite Cache ID 6MLqIQS7y]
5. Google. 2014. The Glass Explorer Program URL: http://www.google.com/glass/start/how-to-get-one/ [accessed 2014-01-03] [WebCite Cache ID 6MLqWI0L6]
6. Nosta J. Forbes. 2013 Jun 21. Inside the operating room with Google Glass URL: http://www.forbes.com/sites/johnnosta/2013/06/21/google-glass-in-the-operating-room/ [accessed 2014-01-03] [WebCite Cache ID 6MLqqbxon]
7. Medical Xpress. 2013 Aug 27. First US surgery transmitted live via Google Glass (w/ video) URL: http://medicalxpress.com/news/2013-08-surgery-transmitted-google-glass-video.html [accessed 2014-01-03] [WebCite Cache ID 6MLrCrQI9]
8. Wikipedia. 2014. Google Glass URL: http://en.wikipedia.org/wiki/Google_Glass [accessed 2014-01-03] [WebCite Cache ID 6MLrU9nbc]
9. Engelen L. LinkedIn. 2013 Aug 15. Is Google Glass useful in the operating room? URL: http://www.linkedin.com/today/post/article/20130815203138-19886490-google-glas-in-or?trk=mp-author-card [accessed 2014-01-03] [WebCite Cache ID 6MLrhZRnG]
10. Glauser W. Doctors among early adopters of Google Glass. CMAJ 2013 Nov 5;185(16):1385. [doi: 10.1503/cmaj.109-4607] [Medline: 24082019]
11. Fox BI, Felkey BG. Potential uses of google glass in the pharmacy. Hosp Pharm 2013 Oct;48(9):783-784 [FREE Full text] [doi: 10.1310/hpj4809-783] [Medline: 24421554]
12. Bankl H. Kapitel 8: Das Obduktionsprotokoll. In: Arbeitsbuch Pathologie, Einführung in die Pathologie - Pathologisch-anatomisches Praktikum. Wien: Facultas Universitätsverlag; 1998:30.
13. Peterson GF, Clark SC, National Association of Medical Examiners. Forensic autopsy performance standards. Am J Forensic Med Pathol 2006 Sep;27(3):200-225. [doi: 10.1097/01.paf.0000243580.43150.3c] [Medline: 16936503]
14. Schneider V, Rothschild MA. Chapter 1.4. In: Brinkmann B, Madea B, editors. Handbuch gerichtliche Medizin. Berlin, Heidelberg: Springer Medizin; 2004.
15. Virchow R. Die Sektions-Technik im Leichenhause des Charité-Krankenhauses, mit besonderer Rücksicht auf gerichtsärztliche Praxis. Berlin: Hirschwald; 1893.
16. Camps FE. Gradwohl's Legal Medicine. Bristol: Wright & Sons; 1976.
17. Knight B. The forensic autopsy. In: Knight B, editor. Forensic Pathology. London: Arnold; 1996:1-49.
18. Council of Europe Committee of Ministers. 1999 Feb 02. Recommendation No R (99) 3 of the Committee of Ministers to Member States on the Harmonisation of Medico-Legal Autopsy Rules URL: http://www.coe.int/T/DG3/Healthbioethic/Texts_and_documents/RecR%2899%293.pdf [accessed 2014-01-04] [WebCite Cache ID 6MN6zchYK]
19. Klein A. Chapter 1.8: Sektionstechnik. In: Brinkman B, Madea B, editors. Handbuch gerichtliche Medizin. Berlin, Heidelberg: Springer Medizin; 2004.
20. Google. Google Glass: Frequently Asked Questions-Tech Specs URL: https://support.google.com/glass/answer/3064128?hl=en&ref_topic=3063354 [accessed 2014-01-03] [WebCite Cache ID 6MLscCTn0]
21. The Gimp Documentation Team. GNU Image Manipulation Program User Manual. Chapter 8: The Colors Menu URL: http://docs.gimp.org/en/gimp-layer-white-balance.html [accessed 2014-01-03] [WebCite Cache ID 6MLrss1f4]
22. Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bulletin 1945;1(6):80-83.
23. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 1947 Mar;18(1):50-60. [doi: 10.1214/aoms/1177730491]
24. Koch GG. Intraclass correlation coefficient. In: Kotz S, Johnson NL, editors. Encyclopedia of Statistical Sciences, 4th edition. New York: John Wiley & Sons; 1982:213-217.
25. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
26. Brooke J. SUS: A “quick and dirty” usability scale. In: Jordan PW, Weerdmeester B, Thomas A, McClelland IL, editors. Usability Evaluation in Industry. London: Taylor & Francis; 1996:107-114.

27. Hassenzahl M, Burmester M, Koller F. Der User Experience (UX) auf der Spur: Zum Einsatz von www.attrakdiff.de. In: Brau H, Diefenbach S, Hassnezahl M, Koller F, Peissner M, Röse K, editors. Usability Professionals. Stuttgart: German Chapter der Usability Professionals Association; 2008:78-82.
28. Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff. Rockville, MD: US Food and Drug Administration; 2013 Sep 25. URL: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf [accessed 2014-01-03] [WebCite Cache ID 6MLs7ynFS]
29. Albrecht UV. Transparency of health-apps for trust and decision making. J Med Internet Res 2013 Dec;15(12):e277 [FREE Full text] [doi: 10.2196/jmir.2981] [Medline: 24449711]
30. Honan M. Wired. 2013 Dec. I Glasshole: My Year With Google Glass URL: http://www.wired.com/gadgetlab/2013/12/glasshole/ [accessed 2014-01-03] [WebCite Cache ID 6MLsOJ5jc]
31. Albrecht UV, von Jan U, Sedlacek L, Groos S, Suerbaum S, Vonberg RP. Standardized, App-based disinfection of iPads in a clinical and nonclinical setting: comparative analysis. J Med Internet Res 2013;15(8):e176 [FREE Full text] [doi: 10.2196/jmir.2643] [Medline: 23945468]
32. Vallurupalli S, Paydak H, Agarwal SK, Agrawal M, Assad-Kottner C. Wearable technology to improve education and patient outcomes in a cardiology fellowship program - a feasibility study. Health Technol 2013 Oct 11;3(4):267-270. [doi: 10.1007/s12553-013-0065-4]
33. Parslow GR. Commentary: Google Glass: A head-up display to facilitate teaching and learning. Biochem Mol Biol Educ 2014;42:91-92. [doi: 10.1002/bmb.20751] [Medline: 24343982]
34. Dunn J. Edudemic. 2013. The teacher's guide to Google Glass URL: http://www.edudemic.com/guides/the-teachers-guide-to-google-glass/ [accessed 2014-01-03] [WebCite Cache ID 6MLzFMeqc]
35. Albrecht UV, Folta-Schoofs K, Behrends M, von Jan U. Effects of mobile augmented reality learning compared to textbook learning on medical students: randomized controlled pilot study. J Med Internet Res 2013;15(8):e182 [FREE Full text] [doi: 10.2196/jmir.2497] [Medline: 23963306]

Abbreviations

API: application programming interface
DSLR: digital single-lens reflex
Exif: exchangeable image file format
ICC: intraclass correlation
IQR: interquartile range
Q1: first quartile
Q2: second quartile (median)
Q3: third quartile

Edited by G Eysenbach; submitted 05.01.14; peer-reviewed by J Nosta, J Tully, C Dameff, T Aungst; comments to author 27.01.14; revised version received 31.01.14; accepted 02.02.14; published 12.02.14.

Please cite as:
Albrecht UV, von Jan U, Kuebler J, Zoeller C, Lacher M, Muensterer OJ, Ettinger M, Klintschar M, Hagemeier L
Google Glass for Documentation of Medical Findings: Evaluation in Forensic Medicine
J Med Internet Res 2014;16(2):e53
URL: http://www.jmir.org/2014/2/e53/
doi:10.2196/jmir.3225
PMID:24521935

Original Paper

Smoking Cessation Apps for Smartphones: Content Analysis With the Self-Determination Theory

Jounghwa Choi[1], PhD; Ghee-Young Noh[1], PhD; Dong-Jin Park[1], PhD

Hallym University, Chuncheon, Korea, Republic Of

Corresponding Author:
Jounghwa Choi, PhD
Hallym University
13406 Damheon Hall, 39 Hallymdaehak-gil
Chuncheon,
Korea, Republic Of
Phone: 82 33 248 1928
Fax: 82 33 256 3424
Email: jhchoi@hallym.ac.kr

Abstract

Background: Smartphones are increasingly receiving attention from public health scholars and practitioners as a means to assist individuals' health management. A number of smartphone apps for smoking cessation are also available; however, little effort has been made to evaluate the content and functions of these apps employing a theoretical framework.

Objective: The present study aims to analyze and evaluate the contents of smoking cessation apps available in South Korea employing the self-determination theory (SDT) as a theoretical framework for analysis. This study analyzes the extent to which smoking cessation apps have features that satisfy the basic needs identified in the SDT, which stimulate autonomous motivation. The type of motivational goal content manifested in the apps and how the goal content was framed are also explored. By assessing the features of smoking cessation apps based on the SDT, this study aims to offer direction for improvement for these apps.

Methods: Out of 309 apps identified from the iTunes store and Google Play (excluding 27 duplications), 175 apps were randomly drawn and analyzed. The coding scheme was drafted by the authors based on the SDT and gain/loss framing theory and was further finely tuned through the process of coder training and by establishing intercoder reliability. Once the intercoder reliability was established, the coders divided up the rest of the sample and coded them independently.

Results: The analysis revealed that most apps (94.3%, 165/175) had at least one feature that tapped at least 1 of the 3 basic needs. Only 18 of 175 apps (10.3%) addressed all 3 basic needs. For goal content, money (53.7%, 94/175) showed the highest frequency, followed by health (32.0%, 56/175), time (7.4%, 13/175), and appearance (1.1%, 2/175), suggesting that extrinsic goals are more dominantly presented in smoking cessation apps. For the framing of goal content, gain framing appeared more frequently (41.7%, 73/175).

Conclusions: The results suggest that these smoking cessation apps may not sufficiently stimulate autonomous motivation; a small number of apps addressed all 3 basic needs suggested by the SDT (ie, autonomy, competence, and relatedness). The apps also tended to present extrinsic goal content (primarily in terms of money) over intrinsic ones (ie, health) by primarily adopting gain framing. Implications of these findings for public health practitioners and consumers are discussed.

(J Med Internet Res 2014;16(2):e44) doi:10.2196/jmir.3061

KEYWORDS

medical informatics applications; smoking cessation; communication media

Introduction

Background

New information technologies are being used to assist continuous health behavior change and the management of chronic disease [1,2]. Smartphones are receiving increasing attention as a potential information technology. Because of the portability of smartphones, which enables people access to them 24 hours a day, long-term management and reinforcement of health behaviors through a variety of communications and apps becomes possible. Based on the recognition of this potential, a variety of smartphone apps for smoking cessation have been released. The problem, however, is that apps developed by individuals are being distributed extensively and it is hard to see if these apps were actually developed based on theoretical and scientific evidence. It is possible that apps with unverified contents have an adverse effect on the health of citizens. For example, a news report argues that many smartphone apps for depression could actually produce countereffects by fostering depression rather than offering a cure for it [3]. Only a few academic studies address this issue [4-6], and little effort has been made in evaluating the contents and functions of the apps employing theoretical framework.

To address this issue further, the present study aimed to analyze and evaluate the contents of smoking cessation apps. We used the self-determination theory (SDT) [7,8] as a theoretical framework for analysis. Although popular health behavior theories, such as the health belief model [9] or the theory of reasoned action [10], are limited in addressing the mechanism underlying the maintenance of changed behaviors [11], SDT offers an insight into this issue, which is critical to smoking cessation [12,13]. According to SDT, the types of motivation that drive individuals' behaviors can vary on the degree of autonomy; if behaviors are triggered by autonomous motivation rather than controlled motivation, people are more likely to engage voluntarily in behavioral changes, which in turn are more likely to be sustained. A number of empirical studies support the effect of autonomous motivation on long-term behavior change, which lasts for a year or longer [14-17]. Therefore, smoking cessation apps with features and contents that are configured by SDT are expected to be more effective.

We analyzed the extent to which smoking cessation apps have features that satisfy the basic needs identified in SDT, which in turn stimulate autonomous motivation. The types of motivational goal contents manifested in the apps and how the goal contents are framed were also explored. By assessing the features of smoking cessation apps based on SDT, this study offers direction for improvement for these apps.

Self-Determination Theory

Overview

According to SDT, gratification of 3 basic psychological needs—autonomy, competence, and relatedness—is essential to the development of intrinsic or autonomous motivation and the maintenance of behavioral change [7,18]. The following discussion will examine how these basic psychological needs could be applied to health behaviors, especially to smoking cessation behavior.

Autonomy

The need for autonomy refers to the individual's need for regulating their behaviors based on their own values, interests, or satisfaction for a behavior itself, not by external influences [18]. SDT research has proposed that social environmental factors, such as autonomy support, play an important role in inducing autonomous motivation by satisfying one's need for autonomy. The effects of autonomy support are evidenced in smoking-related research. According to studies by Williams et al [15,19], an autonomy-supportive communication style induced more autonomous motivation for smoking cessation than a non-autonomy-supportive one. This effect is not only for the short term, but also is likely to last for long-term smoking cessation.

Autonomy support is characterized by the following: (1) providing meaningful reasons for a behavior, (2) providing choices and alternatives for a behavior, (3) supporting individuals' initiatives, and (4) acknowledging individuals' perspectives (eg, negative affect regarding difficult behaviors) [20,21]. To this regard, we can expect that smoking cessation apps with such characteristics would bear greater effect on smoking cessation.

Competence

The need for competence refers to the individual's need for feeling competent about themselves and improving their skills or talents [18]. SDT explains that experience of self-confidence and capabilities related to a behavioral change is necessary for autonomous motivation. The positive effect of competence has been found in studies on smoking cessation [21,22]. These studies showed that competence, along with autonomy, increases the probability of maintaining long-term smoking cessation. These findings are in-line with previous studies, which have reported that self-efficacy or perceived control plays an important role in smoking cessation [23,24].

A sense of competence would be formed and supported by skills, tools, or pertinent feedback that assist individuals to implement behavioral changes as well as to overcome obstacles that inhibit behavioral changes [25]. Therefore, smartphone apps would promote competence for smoking cessation with features such as: (1) providing informational resources concerning skills and knowledge for smoking cessation, (2) providing supporting tools for behavioral implementation, and (3) providing feedback on behaviors implemented and on the individual's progress.

Relatedness

The need for relatedness concerns feelings connected to others [18]. According to SDT, relatedness is important for autonomous motivation because it promotes the internalization of extrinsic causes; individuals may voluntarily engage in behavioral change because of someone else who means a lot to them, even though the behavior by itself is not interesting. In fact, the importance of relatedness, noted as "social support," has been evidenced in various domains of health behaviors, such as breast cancer [26], human immunodeficiency virus (HIV) [27], and exercise [28]. The positive effect of social support on smoking cessation has also been reported in several studies [29-31].

Social support is now available in cyberspace owing to advancements in information technology. For example, various forms of online communities, such as chat rooms, weblogs, or bulletin board systems, can provide social support related to health behaviors [26,32]. The up-and-coming social media is also expected to contribute to social capital formation by enabling individuals to connect with one another [33]. Smartphone apps are often developed in conjunction with social media or offer functions to access communities online. By taking advantage of these characteristics, smoking cessation apps are expected to satisfy relatedness.

In summary, autonomy, competence, and relatedness are important factors to induce intrinsic motivation for long-term behavioral changes. Smoking cessation apps addressing these basic needs are expected to have greater effects on changing smoking behavior. To this regard, this study proposes the first research question: To what extent do smoking cessation apps have features that contribute to the satisfaction of the basic needs (ie, autonomy, relatedness, and competence)?

Goal Contents and Smoking Cessation Behavior

According to SDT, the way individuals select and internalize goals for their own life affects the individual's behavior and psychological well-being. SDT distinguishes intrinsic goal contents (eg, personal growth, health, affiliation) from extrinsic goal contents (eg, financial success, fame, physical attractiveness) [34]. Previous research suggested that intrinsic goals, compared to extrinsic goals, have a more positive and long-term impact on behavioral changes by activating autonomous motivation [35-37]. The advantage of intrinsic goal contents remains significant irrespective of individuals perceiving an activity as instrumental for intrinsic or extrinsic goal attainment [38]; even when individuals pursue extrinsic goals for autonomous reasons, the negative effect of extrinsic goal aspiration remains significant [18]. Regarding smoking cessation, studies have indicated that personalized feedback works better than financial incentive to increase the cessation rates and to prevent relapse [39]; the higher the extent to which one craves physical health as a goal, the greater the readiness to quit smoking or the higher the rate of maintaining a nonsmoking status compared to pursuing cessation for extrinsic reasons (eg, immediate reinforcement such as saving money) [40-42] and those who receive monetary compensation are less likely to remain nonsmokers for long compared to those without monetary compensation [43]. For this reason, Ryan and Deci [8] expressed concern about the use of controlling mechanisms, such as monetary benefits, to promote behavioral changes.

Given the relationship between goal content and long-term behavioral change, it is important to examine what type of goal content is suggested in smoking cessation apps. In smoking cessation apps, the goal content can be implied by presenting the consequences of behaviors. For example, many smoking cessation apps provide a function that calculates monetary gain or loss from smoking or not smoking cigarettes, whereas some provide a function that calculates health consequences, such as life time lost from smoking or blood concentration of carbon monoxide. The former can be viewed as presenting extrinsic goals in terms of the financial aspect, whereas the latter present intrinsic goals in terms of health. Therefore, this study proposes the second research question: To what extent is each type of goal content (ie, intrinsic vs extrinsic) presented in smoking cessation apps?

The Gain/Loss Framing of Goal Contents and Smoking Cessation Behavior

Goal content can be presented differently depending on the type of framing employed. Accordingly, the impact of goal content may rely not only on the kind of goal, but also on the kind of framing employed to present the goal. Gain framing emphasizes the positive outcomes, either physical or psychological benefits, which will be gained from adopting the recommendations. Loss framing obtains compliance by emphasizing the negative consequences resulting from not adopting the recommendations. Pelletier and Sharp [44] argued that both intrinsic and extrinsic goals could be presented in a gain or loss frame. For example, a message may suggest that one can gain health by engaging in a behavior (ie, gain framing of an intrinsic goal) or that one can lose money by engaging in a behavior (ie, loss framing of an extrinsic goal).

It has been proposed that gain framing works better for prevention behaviors (eg, smoking cessation) and loss framing works better for detection behaviors (eg, mammography) [45]. Although some scholars argue for the meager effect of a gain-framed message for prevention behavior [46], gain framing's advantage over loss framing for prevention behaviors is relatively robust [47-48]. Recently accumulated evidence, including Gallagher and Updegraff's meta-analysis [49], provide support for this argument, at least for smoking cessation [50-54]. Given the evidence for the effectiveness of smoking cessation apps adopting gain-framed information, the third research question is put forth to assess the extent to which the apps adopt gain framing over loss framing in presenting the goal content: To what extent are gain and loss framing used in presenting the goal content implied in smoking cessation apps?

Methods

Sample

In this study, smoking cessation apps are defined as apps that aim to induce behavioral changes to quit smoking. Although app markets are not strictly bounded by region, an app available in one country may not be available in other countries. Therefore, we limited our analysis to the apps available in South Korea, which the authors had access to. To draw a sample, apps in Google Play and the Apple iTunes store, the largest open markets for smartphone apps, were searched during the last week of November 2013 utilizing the keywords “smoking” and “smoking cessation,” either in Korean or in English. Although the search from the iTunes store produced a definite number of cases for each keyword, the total number was not identifiable for Google Play (ie, expressed as +100,000). Therefore, for Google Play, the initial search result was sorted by relevance and then the Web pages were reviewed until a Web page without a smoking cessation app was found. A list of apps was compiled by examining the title and the description of apps searched. That is, for both the iTunes store and Google Play, we used an additive strategy (ie, adding apps to the list while reviewing the resulting search pages for each keyword), rather than a deductive strategy (ie, deducting the irrelevant apps from the definite number apps from the initial search) to compile the list of apps.

Among the apps searched with keywords, those which were not relevant to smoking cessation were excluded. Apps were also excluded even if they had some relevance to smoking cessation in the following cases: (1) task management apps were dropped unless their primary purpose was to aid smoking cessation; (2) hypnosis apps for smoking cessation were disregarded because they attempt to exert a subconscious influence and are not appropriate to be analyzed within the frame of SDT; (3) apps developed for physicians to aid their medical treatment, rather than for general consumers, were also not included; (4) apps offering simulation of smoking were also not included unless they clearly stated their purpose as smoking cessation. As a result, 309 apps were included: 167 from the iTunes store and 169 from Google Play, with 27 duplications identified. To draw a sample, 175 apps were randomly drawn based on sample size guidelines [55,56]. The apps searched and downloaded from December 2012 through January 2013 were analyzed.

Coding of Apps

Coding Scheme

The draft of the coding scheme was developed based on the aforementioned conceptualization of each variable, driven by SDT and the gain/loss framing theory. Based on this draft, the authors reviewed a part of the smoking cessation apps and revised the draft, taking into account their characteristics. This version of the draft was further fine-tuned through the process of coder training and establishing intercoder reliability.

Descriptive Characteristics

To identify the general characteristics of the smoking cessation apps, the following items were coded: market type (Google Play, iTunes store, or available in both), price type (free vs paid), developer type (individuals or individual developer groups, nonprofit organizations, or companies), and contents type (information-centric, function-centric, or information-function balanced). The number of downloads was also included in the coding scheme; however, this data was collected for Android apps only because the iTunes store does not offer this information.

Autonomy

To assess autonomy, we examined whether an app had features related to the aforementioned autonomy-supportive characteristics. First, apps were examined to determine if they offered relevant information addressing reasons for stopping smoking (eg, scientific evidence on the risk of smoking). Apps were coded as "yes" on this if they offered a separate section for the information, otherwise they were coded as "no." That is, apps presenting a few sentences simply stating, “smoking is bad for health,” were coded as "no" on this. Second, apps were coded as "yes" if they offered functions allowing one’s own smoking cessation plan (eg, setting up a cessation schedule or amount of cigarettes to smoke; choosing cessation methods) to support the individual’s initiatives, the second aspect of autonomy support. These functions were also deemed as tapping the third aspect of autonomy support (ie, providing individuals choices and alternatives for smoking cessation behavior) because they allow individuals with different approaches to quit smoking using their own time frame and strategies. The last aspect of autonomy is acknowledging individuals’ perspectives or negative affects experienced from quitting smoking. This only can be assessed in the context of interpersonal communication; therefore, apps were not coded on this.

Competence

Competence is formed and supported by skills, tools, or pertinent feedback that assist individuals to implement behavioral changes [25]. Therefore, features related to competence were assessed in light of the absence (coded as "no") or presence (coded as "yes") of the following features in an app. First, apps were coded as "yes" if they offered information resources concerning skills and knowledge for smoking cessation; this included offering specific guidelines or tips for course of actions to quit smoking and resources to increase knowledge regarding smoking cessation (eg, FAQ, hotline, links to website, quiz or game to improve knowledge related to smoking cessation). Second, the absence or presence of tools assisting the implementation of behavioral change was assessed, which included functions to send alert or alarm messages to warn or remind, and functions to record and track one’s own quitting or smoking attempts. Finally, apps were coded as "yes" if they offered feedback functions, such as offering an analysis of one’s own quitting attempts or efforts or providing cues to progress or achievement toward smoking cessation.

Relatedness

Because relatedness concerns feelings connected to others, the apps were examined in terms of whether they offered functions that allowed for interacting with others or receiving social support from others. Specifically, the presence (coded as "yes") or absence (coded as "no") of the following attributes were coded: providing functions to interact with others (eg, online communities, social media), and providing functions to deliver social support messages (eg, recording and playing messages from family members or friends).

Goal Contents and Framing

To analyze the goal content and the gain/loss framing implied in the smoking cessation apps, we examined the consequences of quitting smoking that manifested in the functions of smoking cessation apps based on the initial identification by the authors. Specifically, the absence (coded as "no") or presence (coded as "yes") of each of the 4 types of goal content (eg, money, appearance, health, and time) were coded because apps often presented multiple goal content simultaneously. Each type of goal content identified was further examined to see if it was presented in terms of gain or loss. For example, when a function was identified as presenting money as a goal, it was examined to determine if it concerned calculating monetary loss from smoking or gain from quitting smoking. It was possible that an app could offer both gain-framed and loss-framed functions; thus, the presence (coded as "yes") or absence (coded as "no") of each function was coded. Gain/loss framing of other goal content type was also measured in the same manner: appearance (ie, a function that shows appearance deterioration from smoking or appearance improvement from quitting smoking), health (ie, a function that shows health-related loss from smoking or health-related gain from quitting smoking, such as life expectancy or blood pressure), and time (ie, a function that calculates time wasted from smoking/saved from quitting smoking). Money and appearance are deemed extrinsic goals and health is considered an intrinsic goal [18]. In all, 7.5% (13/175) of apps suggested time wasted or saved as a consequence of smoking or quitting smoking. However, no theoretical rationale was found for the category of time; thus, we did not categorize it into either category.

Results

Intercoder Reliability

Intercoder reliability was established following the guidelines by Lombard et al [57]. Two undergraduate students fluent in both English and Korean served as coders. They became acquainted with the coding scheme through 2 training sessions. Examining the apps based on the coding scheme, the coders were asked to try every component in an app. For example, when apps required users to enter their own information, create an account, or take a test, the coders did so to fully appreciate the features of the apps. It took the coders approximately 15 minutes to code a simple app and 30-40 minutes for a more complex app. For initial intercoder reliability, the coders independently coded 30 apps that were not part of the sample. Intercoder reliability was calculated with PRAM ver 0.4.5 [58]. For some items, the reliability was less than .70; thus, some of the initial coding categories were adjusted after discussion between the researchers and the coders through another training session and the coding scheme was finalized as described previously. For final intercoder reliability, 50 apps (28.6% of the total sample) were drawn from the sample. The percent agreement was greater than .90 for all coding items. The Cohen's kappa or Scott's pi coefficients were >.70 for most coding items. However, the coefficients were zero for loss framing of time and 2 attributes of relatedness (ie, connected to online communities, connected to social media). The coefficient for loss framing of appearance turned out to be negative. This could happen for binary variables in which 1 of the values (1 is present or 0 is absent) is observed very rarely or not at all because these coefficients take into account the prevalence of the categories [59]. An examination of the data revealed that no case in the sample data for intercoder reliability was coded on loss framing of time and 2 attributes of relatedness. As for the loss framing of appearance, both coders agreed the cases were absent of this feature, except for 1 case. In the process of resolving the difference, it was found that 1 coder missed the feature while reviewing the app, as opposed to judging the feature differently from the other coder. Because the percent agreement for those items exceeded .90 and it was possible for more cases to be observed in the whole sample, we decided to retain these items. The disagreement between the coders was resolved through a discussion with both coders and researchers. After establishing intercoder reliability, the coders divided up the rest of the sample and coded them independently.

Descriptive Statistics of the Sample

The descriptive characteristics of the apps were examined in terms of market, price, developer, and content type (Table 1). Among 175 apps, 14 (8.0%) were found both in Google Play and the iTunes store. The rest were solely from either the iTunes store (45.7%, 80/175) or Google Play (46.3%, 81/175). Overall, free apps (55.4%, 97/175) were more frequently observed than paid apps (44.6%, 78/175). Most apps were developed by individual developers or developer groups (68.0%, 119/175). Only 9 apps (5.1%) were found to be developed by nonprofit organizations. For content type, most apps tended to be function oriented (60.0%, 105/175). Approximately one-fifth (18.9%, 33/175) of the apps were judged to balance function with information.

Table 1. Descriptive statistics of smoking cessation apps in the sample (N=175).

Analysis categories	Distribution, n (%)
Market type	
Android	81 (46.3)
Apple	80 (45.7)
Android/Apple	14 (8.0)
Price type	
Free	97 (55.4)
Paid	78 (44.6)
Developer type	
Individual or developer group	119 (68.0)
Nonprofit organization	9 (5.1)
Company	47 (26.9)
Contents type	
Information oriented	37 (21.1)
Function oriented	105 (60.0)
Information and function balanced	33 (18.9)

Features of Smoking Cessation Apps Related to Basic Needs

The first research question asked the extent to which smoking cessation apps have features related to the basic needs (ie, autonomy, relatedness, and competence). Our analysis of the coded data of the apps revealed that most apps (94.3%, 165/175) had at least one feature which tapped at least 1 of the 3 basic needs, whereas few apps (10.3%, 18/175) addressed all 3 basic needs. Calculation of the total basic need score by summing the scores on all 3 basic needs produced a mean of 2.82 (SD 1.56), indicating that smoking cessation apps have approximately 3 kinds of features tapping any of the basic needs.

At least one feature related to competence was found in most apps (86.9%, 152/175). These features included offering how-to information regarding smoking cessation (40.0%, 70/175), functions to record and track one's own quitting or smoking attempts (34.3%, 60/175), or cues to achievement or progress (44.0%, 77/175). In regards to autonomy, approximately half (48.6%, 85/175) of the apps contained at least one feature supporting autonomy. The observed features tapping autonomy were information regarding reason for smoking cessation (33.7%, 59/175) and functions allowing one's own smoking cessation plan (20.0%, 35/175). In contrast to competence and autonomy, a smaller proportion of the apps addressed relatedness (21.1%, 37/175). The most frequently observed feature related to relatedness was connection to social media (17.7%, 31/175), through which the users could share their own efforts to quit smoking and receive supportive messages from their acquaintances (Table 2).

Table 2. Analysis of smoking cessation app features related to basic needs (N=175).

Basic needs and app features	n (%)
Autonomy	
Information addressing reason for smoking cessation	59 (33.7)
Function allowing one's own smoking cessation plan	35 (20.0)
Offers at least 1 of the autonomy features above	85 (48.6)
Competence	
How-to information and resources	70 (40.0)
Tools assisting implementation of behavior	
Functions to send alert or alarm messages to warn or remind	18 (10.3)
Function to record and track one's own quitting or smoking attempts	60 (34.3)
Offers feedback	
Analysis of performance for quitting attempts or efforts	50 (28.6)
Cues to achievement or progress	77 (44.0)
Offers at least 1 of the competence features above	152 (86.9)
Relatedness	
Function to interact with others	
Connected to online communities	7 (2.4)
Connected to social media	31 (17.7)
Function to offer or record messages to cheer up	2 (1.1)
Offers at least 1 of the relatedness features above	37 (21.1)

Goal Content and Framing in Smoking Cessation Apps

The second research question attempted to address the extent to which each type of goal content (ie, intrinsic vs extrinsic) was implied in the smoking cessation apps. Among the 4 types of goal content identified in the coding scheme, money (53.7%, 94/175) showed the highest frequency, followed by health (32.0%, 56/175), time (7.4%, 13/175), and appearance (1.1%, 2/175) (Table 3). When grouping money and appearance as extrinsic goals and health as an intrinsic goal, 29.1% (51/175) of the apps were found to present both extrinsic and intrinsic goals simultaneously. The apps suggesting only extrinsic goals were 25.7% (45/175), whereas the apps suggesting only intrinsic goals were minimal (2.9%, 5/175) (Table 4). This result suggests that extrinsic goals are more dominantly presented among smoking cessation apps than intrinsic goals.

The third research question concerned the framing of goal content implied in the smoking cessation apps. Analysis of the data revealed a more frequent appearance of gain framing. That is, the apps tended to focus on monetary gain (ie, amount of money saved from not smoking a cigarette) or health benefits (eg, life time earned or blood pressure decreased from quitting smoking) rather than monetary loss (ie, amount of money wasted from smoking a cigarette) or health loss (eg, life time lost or blood pressure increased from smoking) (Table 3).

Table 3. Implied goal content and framing type in smoking cessation apps available in South Korea (N=175).

Goal content and framing	n (%)
Money	94 (53.7)
Gain only	57 (32.6)
Loss only	25 (14.3)
Gain and loss both	12 (6.9)
Health	56 (32.0)
Gain only	38 (21.7)
Loss only	14 (8.0)
Gain and loss both	4 (2.3)
Time	13 (7.4)
Gain only	8 (4.6)
Loss only	2 (1.1)
Gain and loss both	3 (1.7)
Appearance	2 (1.1)
Gain only	0 (0.0)
Loss only	2 (1.1)
Gain and loss both	0 (0.0)

Table 4. Goal content (intrinsic vs extrinsic) of smoking cessation apps available in South Korea (N=175).

Category	n (%)
Intrinsic goal content (ie, health) only	5 (2.9)
Extrinsic goal content (ie, money and appearance) only	45 (25.7)
Intrinsic and extrinsic goal content both	51 (29.1)
Total number of cases suggesting goal content	101 (57.7)

Based on this analysis, we identified the top smoking cessation apps. First, we identified the apps that had at least one feature that tapped each of the 3 basic needs. In all, 18 apps met this criterion. These were then sorted by total SDT score, the sum of scores for each SDT attribute. Among the top-ranked apps, 4 apps developed in the public sectors are presented in Table 5 along with their unique features: Singapore's Health Promotion Board (Figure 1), Tobacco Control Research Branch at the National Cancer Institute in the United States (Figure 2), the European Commission (Figure 3), and the Korean Ministry of Health and Welfare (Figure 4). They tended to contain common features, such as aiding one's own quitting plan and management of it, providing tips and strategies, and allowing sharing through social media.

Table 5. Top-ranked smoking cessation apps developed in the public sector.

Name of app	Rank[a]	Market type	Unique features
HPB I Quit [60]	1	Apple/Android	Developed by the Singapore Health Promotion Board. Links to QuitLine and quit centers. Introduces various methods to quit smoking.
QuitSTART [61]	3	Apple	Developed by Tobacco Control Research Branch at the US National Cancer Institute. Offers symbolic signs of achievement; tracking, and analysis of behaviors.
FCB/Exsmokers iCoach [62]	3	Apple/Android	Developed by the European Commission. Aid quitting plan and managing it step-by-step; offers analysis of behaviors.
[No Smoke Guide][b] [63]	7	Apple/Android	Offered by the Korean Ministry of Health and Welfare. Offers information on quit centers; in-depth information on smoking and health.

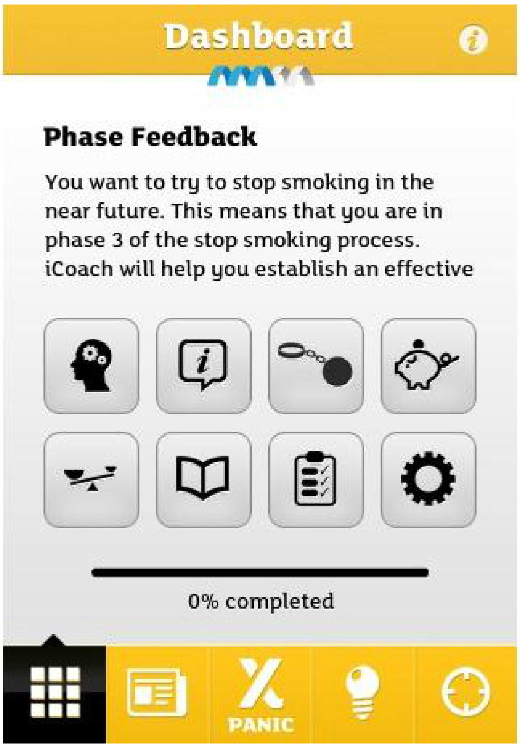
[a] Rank score from top 10 list of apps that included paid and free apps.

[b] App in Korean (name of app translated into English).

Figure 1. Screenshot of the HPB I Quit app by the Singapore Health Promotion Board.

Figure 2. Screenshot of the QuitSTART app by the US Tobacco Control Research Branch at the National Cancer Institute.

Figure 3. Screenshot of the FCB/Exsmokers iCoach app by the European Commission.



Phase Feedback

You want to try to stop smoking in the near future. This means that you are in phase 3 of the stop smoking process. iCoach will help you establish an effective

Figure 4. Screenshot of the [No Smoke Guide] app by the Korean Ministry of Health and Welfare.

Discussion

Principal Results

Health promotion through personal media, such as smartphones, is receiving increasing attention as a variety of health-related smartphone apps are being introduced in the market. However, only a few studies have evaluated the content and function of these apps by employing a theoretical framework. Therefore, the present study analyzed a representative sample of smoking cessation apps (N=175) accessible in South Korea based on the theoretical framework of SDT.

The analysis suggests that most apps contain at least 1 kind of feature related to at least 1 of the components of the basic needs proposed by SDT, primarily competence. A relatively small proportion of the apps, however, addressed all 3 basic needs (18/175, 10.3%). In particular, only a few apps supporting relatedness were found (37/175, 21.1%). SDT studies have shown that satisfaction of the 3 basic psychological needs is fundamental to meaningful and continuous behavioral change [7], and this is also supported by research in smoking cessation [15,21]. It might be reasonable to say that smoking cessation apps that lack 1 or more components of the basic needs may be relatively limited in producing positive effects on continuous behavioral changes compared to those satisfying all aspects of basic needs. The efficacy of the apps would depend on how individuals make use of them. Notwithstanding, it is expected that there is at least more potential that the apps with features addressing the 3 basic needs could be more efficacious.

The analysis of goal content implied in the functions of smoking cessation apps also reveals interesting findings. The most frequently observed type of goal content was money (an extrinsic goal) followed by health (an intrinsic goal). More than half of the smoking cessation apps (53.7%, 94/175) offered a function that calculated money earned or lost per cigarette. In addition, the smoking cessation apps tended to include either extrinsic goal content only or extrinsic and intrinsic goal content together. Given that extrinsic goal content is more likely to be related to controlled motivation, and controlled motivation is more likely to be induced even when both extrinsic and intrinsic goal content are suggested at the same time [64], these smoking cessation apps may be limited in activating autonomous motivation. Because behavioral changes prompted by controlled motivation are less likely to last, the efficacy of these smoking cessation apps in inducing long-term behavioral change is open to question.

For the analysis of framing of goal content, it was found that gain framing was more frequently adopted in suggesting goal content; apps tended to offer a function that calculated the financial or health gain from quitting smoking rather than money or health loss from smoking. As mentioned previously, gain framing has been found to be more effective than loss framing for prevention behaviors, such as quitting smoking [49]. Further, gain framing is more effective than loss framing in moving smokers toward the preparation stage from the contemplation stage of smoking cessation [65]. Those who download the smoking cessation apps might possibly be interested in quitting smoking or have decided to quit smoking (ie, those who are in the contemplation or preparation stage). In this sense, the fact that smoking cessation apps are adopting gain framing more than loss framing could be viewed positively.

We expect that the findings of this present study can be generalized to the United States and possibly to other countries as well. Out of the 309 apps that were the population of this study, 21 apps (6.8%) were available only in Korean, not in English. For the remaining apps that were available in English, when we searched the US Google Play and iTunes store with an anonymous helper with a US account, only a few apps (7/309, 2.3%) were not found. It seems that most of these apps are circulated globally, whereas a small portion of the apps may be available only to certain regions. If this is the case, the findings of this study might be sustained in other countries.

Implications and Limitations

The present study is significant in that it conducted a theory-driven analysis using SDT to evaluate smoking cessation apps. SDT explains how individuals internalize external forces, make autonomous decisions, and implement voluntary actions. The SDT explanation of the 3 basic needs and their roles in the motivation process lends a useful theoretical framework to analyze and evaluate smoking cessation apps because the theory informs the kind of characteristics that a smartphone app aiming for long-term behavioral change needs to have to stimulate autonomous motivation. Although this study is not a direct test on the effects of smoking cessation apps, this theory-driven analysis allows an indirect assessment of smoking cessation apps available in the market.

In addition, the present study provides several practical implications for consumers, public health experts, and practitioners. In view of SDT, our study suggests that many of the smoking cessation apps available in the market may be partially limited in inducing long-term smoking cessation. It might be useful for consumers if public health practitioners or organizations evaluate the apps and recommend high-quality smoking cessation apps. Further, providing consumers with a guide regarding how to evaluate smoking cessation apps would be helpful. In addition, public health organizations could develop apps using the evaluation criteria addressed in this study.

The findings of the present study are bounded by several limitations. First, we recognize that our analysis is limited in fully analyzing the whole aspects of an app. For example, a function that is coupled with social media could enhance not only relatedness, but also competence because it is also possible that individuals exchange tips and information for smoking cessation through social media. Also, we primarily focused on the functional features of the apps and did not examine the messages delivered in the apps. Second, some apps had functional features that were not specified on the menu, but activated when used over the longer term. Because our coders did not use the apps for the long term, these features might not have been captured. Third, our study is limited in providing descriptions of all kinds of smoking cessation apps because we excluded hypnosis apps in our analysis. Future studies that examine hypnosis apps in terms of the messages delivered would provide a complementary picture about smoking cessation apps to this study. Finally, the inferences from the findings of this study could be limited because this study is not a direct test of the effects of smoking cessation apps. Rather, it draws an indirect inference regarding their efficacy based on previous SDT studies. It should be noted that the conclusion on the effect of gain or loss framing and the effect of goal content are still open to discussion. Although a substantial number of studies support the advantage of gain-framed messages over loss-framed messages in regards to smoking cessation, researchers also acknowledge that the effect of gain or loss framing could be complex, so that multiple mediators and moderators can play into the framing effect [47-49]. Future studies should address this issue by exploring the effects of smoking cessation apps through experimental studies.

Conclusions

To sum up, the present study shows that smoking cessation apps have features that satisfy the basic needs to some extent; thus, they could be a useful tool to promote smoking cessation. However, it is also true that many of these apps provide limited features to satisfy all 3 basic needs and present extrinsic goals rather than intrinsic goals. As a result, they may serve as a limited tool in stimulating autonomous motivation for long-term smoking cessation. This finding is in-line with studies by Abroms et al [4,5] that concluded that smoking cessation apps, even popular ones, are far below the US Public Health Service's Clinical Practice Guidelines for Treating Tobacco Use and Dependence. Although the present study is only a snapshot of smoking cessation apps presently available, this study suggests that there is still room to increase the efficacy of these apps and that a good theory, such as SDT, can guide the process. Public health practitioners would need to play a role in planning mobile health apps by informing developers of the theory.

Acknowledgments

This study was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012S1A3A2033480).

Conflicts of Interest

None declared.

References

1. Franklin VL, Greene A, Waller A, Greene SA, Pagliari C. Patients' engagement with "Sweet Talk" - a text messaging support system for young people with diabetes. J Med Internet Res 2008;10(2):e20 [FREE Full text] [doi: 10.2196/jmir.962] [Medline: 18653444]
2. Bock B, Heron K, Jennings E, Morrow K, Cobb V, Magee J, et al. A text message delivered smoking cessation intervention: the initial trial of TXT-2-Quit: randomized controlled trial. JMIR Mhealth Uhealth 2013 Jul 30;1(2):e17. [doi: 10.2196/mhealth.2522]
3. Ryu I. Kyunghyang Newspaper. 2012 Feb 12. [Using apps for depression for fun potentially poster depression] URL: http://news.khan.co.kr/kh_news/khan_art_view.html?artid=201202122141395&code=940601 [accessed 2014-01-29] [WebCite Cache ID 6MzX9i5Su]
4. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
5. Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736. [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
6. Muessig KE, Pike EC, Legrand S, Hightow-Weidman LB. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. J Med Internet Res 2013;15(1):e1 [FREE Full text] [doi: 10.2196/jmir.2301] [Medline: 23291245]
7. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
8. Ryan RM, Deci EL. A self-determination theory approach to psychotherapy: The motivational basis for effective change. Canadian Psychol 2008;49(3):186-193. [doi: 10.1037/a0012753]
9. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [Medline: 3378902]
10. Heilbroner RL, Thurow LC, Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. Englewood Cliffs, NJ: Prentice-Hall; 1980.
11. Rothman AJ. Toward a theory-based analysis of behavioral maintenance. Health Psychol 2000 Jan;19(1 Suppl):64-69. [Medline: 10709949]
12. American Cancer Society. 2013. Guide to Quitting Smoking URL: http://www.cancer.org/healthy/stayawayfromtobacco/guidetoquittingsmoking/index [accessed 2013-12-25] [WebCite Cache ID 6M8CasAUc]
13. Zhou X, Nonnemaker J, Sherrill B, Gilsenan AW, Coste F, West R. Attempts to quit smoking and relapse: factors associated with success or failure from the ATTEMPT cohort study. Addict Behav 2009 Apr;34(4):365-373. [doi: 10.1016/j.addbeh.2008.11.013] [Medline: 19097706]
14. Williams GC, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. Diabetes Care 1998 Oct;21(10):1644-1651. [Medline: 9773724]
15. Williams GC, Gagné M, Ryan RM, Deci EL. Facilitating autonomous motivation for smoking cessation. Health Psychol 2002 Jan;21(1):40-50. [Medline: 11846344]
16. Williams GC, Minicucci DS, Kouides RW, Levesque CS, Chirkov VI, Ryan RM, et al. Self-determination, smoking, diet and health. Health Educ Res 2002 Oct;17(5):512-521 [FREE Full text] [Medline: 12408196]

17. Williams GC, Patrick H, Niemiec CP, Ryan RM, Deci EL, Lavigne HM. The smoker's health project: a self-determination theory intervention to facilitate maintenance of tobacco abstinence. Contemp Clin Trials 2011 Jul;32(4):535-543 [FREE Full text] [doi: 10.1016/j.cct.2011.03.002] [Medline: 21382516]
18. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychological Inquiry 2000 Oct;11(4):227-268. [doi: 10.1207/S15327965PLI1104_01]
19. Williams GC, Cox EM, Kouides R, Deci EL. Presenting the facts about smoking to adolescents: effects of an autonomy-supportive style. Arch Pediatr Adolesc Med 1999 Sep;153(9):959-964. [Medline: 10482213]
20. Reeve J, Jang H. What teachers say and do to support students' autonomy during a learning activity. Journal of Educational Psychology 2006;98(1):209-218. [doi: 10.1037/0022-0663.98.1.209]
21. Williams GC, McGregor HA, Sharp D, Levesque C, Kouides RW, Ryan RM, et al. Testing a self-determination theory intervention for motivating tobacco cessation: supporting autonomy and competence in a clinical trial. Health Psychol 2006 Jan;25(1):91-101. [doi: 10.1037/0278-6133.25.1.91] [Medline: 16448302]
22. Williams GC, Niemiec CP, Patrick H, Ryan RM, Deci EL. The importance of supporting autonomy and perceived competence in facilitating long-term tobacco abstinence. Ann Behav Med 2009 Jun;37(3):315-324 [FREE Full text] [doi: 10.1007/s12160-009-9090-y] [Medline: 19373517]
23. Condiotte MM, Lichtenstein E. Self-efficacy and relapse in smoking cessation programs. J Consult Clin Psychol 1981 Oct;49(5):648-658. [Medline: 7287974]
24. Shiffman S, Balabanis MH, Paty JA, Engberg J, Gwaltney CJ, Liu KS, et al. Dynamic effects of self-efficacy on smoking lapse and relapse. Health Psychol 2000 Jul;19(4):315-323. [Medline: 10907649]
25. Ryan RM, Patrick H, Deci EL, Williams GC. The European Health Psychologist. 2008 Mar. Facilitating health behavior change and its maintenance: Interventions based on Self-Determination Theory URL: http://www.ehps.net/ehp/issues/2008/v10iss1_March2008/EHP_March_2008_All.pdf [accessed 2013-12-25] [WebCite Cache ID 6M7khn2Tr]
26. Fogel J, Albert SM, Schnabel F, Ditkoff BA, Neugut AI. Internet use and social support in women with breast cancer. Health Psychol 2002 Jul;21(4):398-404. [Medline: 12090683]
27. Friedland J, Renwick R, McColl M. Coping and social support as determinants of quality of life in HIV/AIDS. AIDS Care 1996 Feb;8(1):15-31. [doi: 10.1080/09540129650125966] [Medline: 8664366]
28. Duncan TE, McAuley E. Social support and efficacy cognitions in exercise adherence: a latent growth curve analysis. J Behav Med 1993 Apr;16(2):199-218. [Medline: 8315646]
29. Coppotelli HC, Orleans CT. Partner support and other determinants of smoking cessation maintenance among women. J Consult Clin Psychol 1985 Aug;53(4):455-460. [Medline: 4031200]
30. May S, West R. Do social support interventions ("buddy systems") aid smoking cessation? A review. Tob Control 2000 Dec;9(4):415-422 [FREE Full text] [Medline: 11106712]
31. 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]
32. Shaw BR, McTavish F, Hawkins R, Gustafson DH, Pingree S. Experiences of women with breast cancer: exchanging social support over the CHESS computer network. J Health Commun 2000;5(2):135-159. [doi: 10.1080/108107300406866] [Medline: 11010346]
33. Ellison NB, Steinfield C, Lampe C. The benefits of Facebook “friends:” social capital and college students’ use of online social network sites. J Comput-Mediat Comm 2007;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
34. Vansteenkiste M, Lens W, Deci EL. Intrinsic versus extrinsic goal contents in self-determination theory: another look at the quality of academic motivation. Educational Psychologist 2006 Mar;41(1):19-31. [doi: 10.1207/s15326985ep4101_4]
35. Gillison FB, Standage M, Skevington SM. Relationships among adolescents' weight perceptions, exercise goals, exercise motivation, quality of life and leisure-time exercise behaviour: a self-determination theory approach. Health Educ Res 2006 Dec;21(6):836-847 [FREE Full text] [doi: 10.1093/her/cyl139] [Medline: 17101718]
36. Sebire SJ, Standage M, Vansteenkiste M. Examining intrinsic versus extrinsic exercise goals: cognitive, affective, and behavioral outcomes. J Sport Exerc Psychol 2009 Apr;31(2):189-210. [Medline: 19454771]
37. Sebire SJ, Standage M, Vansteenkiste M. Predicting objectively assessed physical activity from the content and regulation of exercise goals: evidence for a mediational model. J Sport Exerc Psychol 2011 Apr;33(2):175-197. [Medline: 21558579]
38. Vansteenkiste M, Timmermans T, Lens W, Soenens B, Van den Broeck A. Does extrinsic goal framing enhance extrinsic goal-oriented individuals' learning and performance? An experimental test of the match perspective versus self-determination theory. Journal of Educational Psychology 2008;100(2):387-397. [doi: 10.1037/0022-0663.100.2.387]
39. Curry SJ, Wagner EH, Grothaus LC. Evaluation of intrinsic and extrinsic motivation interventions with a self-help smoking cessation program. J Consult Clin Psychol 1991 Apr;59(2):318-324. [Medline: 2030194]
40. Curry S, Wagner EH, Grothaus LC. Intrinsic and extrinsic motivation for smoking cessation. J Consult Clin Psychol 1990 Jun;58(3):310-316. [Medline: 2195084]
41. Curry SJ, Grothaus L, McBride C. Reasons for quitting: intrinsic and extrinsic motivation for smoking cessation in a population-based sample of smokers. Addict Behav 1997;22(6):727-739. [Medline: 9426790]

42. Niemiec CP, Ryan RM, Deci EL, Williams GC. Aspiring to physical health: the role of aspirations for physical health in facilitating long-term tobacco abstinence. Patient Educ Couns 2009 Feb;74(2):250-257 [FREE Full text] [doi: 10.1016/j.pec.2008.08.015] [Medline: 18838243]
43. Cahill K, Perera R. Competitions and incentives for smoking cessation. Cochrane Database Syst Rev 2011(4):CD004307. [doi: 10.1002/14651858.CD004307.pub4] [Medline: 21491388]
44. Pelletier LG, Sharp E. Persuasive communication and proenvironmental behaviours: How Message framing can improve the integrationof behaviours through self-determined motivation. Canadian Psychol 2008;49:210-217. [doi: 10.1037/a0012755]
45. Rothman AJ, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull 1997 Jan;121(1):3-19. [Medline: 9000890]
46. O'Keefe DJ, Jensen JD. The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease prevention behaviors: a meta-analytic review. J Health Commun 2007;12(7):623-644. [doi: 10.1080/10810730701615198] [Medline: 17934940]
47. Updegraff JA, Rothman AJ. Health message framing: moderators, mediators, and mysteries. Social and Personality Psychology Compass 2013 Sep 02;7(9):668-679. [doi: 10.1111/spc3.12056]
48. Updegraff JA, Rothman AJ, Salovey P. Using Message Framing to Promote Healthful Behaviors: An Update. 2012 Jan 16. URL: http://updegrafflab.org/files/3113/3889/7925/URS-12.pdf [accessed 2013-12-25] [WebCite Cache ID 6M7mApYpm]
49. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
50. Toll BA, O'Malley SS, Katulak NA, Wu R, Dubin JA, Latimer A, et al. Comparing gain- and loss-framed messages for smoking cessation with sustained-release bupropion: a randomized controlled trial. Psychol Addict Behav 2007 Dec;21(4):534-544 [FREE Full text] [doi: 10.1037/0893-164X.21.4.534] [Medline: 18072836]
51. Toll BA, Salovey P, O'Malley SS, Mazure CM, Latimer A, McKee SA. Message framing for smoking cessation: the interaction of risk perceptions and gender. Nicotine Tob Res 2008 Jan;10(1):195-200 [FREE Full text] [doi: 10.1080/14622200701767803] [Medline: 18188760]
52. Szklo AS, Coutinho ES. The influence of smokers' degree of dependence on the effectiveness of message framing for capturing smokers for a Quitline. Addict Behav 2010 Jun;35(6):620-624. [doi: 10.1016/j.addbeh.2010.01.007] [Medline: 20163919]
53. Fucito LM, Latimer AE, Salovey P, Toll BA. Nicotine dependence as a moderator of message framing effects on smoking cessation outcomes. Ann Behav Med 2010 Jun;39(3):311-317 [FREE Full text] [doi: 10.1007/s12160-010-9187-3] [Medline: 20502997]
54. Bresnahan MJ, Zhuang J, Sun S. Influence of smoking norms and gain/loss antismoking messages on young Chinese adults. Nicotine Tob Res 2013 Sep;15(9):1564-1571. [doi: 10.1093/ntr/ntt015] [Medline: 23430738]
55. Krejcie RV, Morgan DW. Determining sample size for research activities. Educ Psychol Meas 1970;30:607-610. [doi: 10.1177/001316447003000308]
56. Bartlett JE, Kotrlik JW, Higgins CC. Information Technology, Learning, and Performance Journal. 2001. Organizational research: Determining appropriate sample size in survey research URL: http://www.osra.org/itlpj/bartlettkotrlikhiggins.pdf [accessed 2013-12-25] [WebCite Cache ID 6M7nBZ9Rp]
57. Lombard M, Snyder-Duch J, Bracken CC. Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. Human Comm Res 2002 Oct;28(4):587-604. [doi: 10.1111/j.1468-2958.2002.tb00826.x]
58. Neuendorf KA. Temporary Online Location for PRAM Downloads. URL: http://academic.csuohio.edu/kneuendorf/c63311/PRAMS/ [accessed 2014-01-29] [WebCite Cache ID 6MyqGEq6L]
59. De Swert K. Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha. 2012. URL: http://www.polcomm.org/wp-content/uploads/ICR01022012.pdf [accessed 2013-10-30] [WebCite Cache ID 6KkG6S9XE]
60. Google Play. Singapore: Health Promotion Board I Quit URL: https://play.google.com/store/apps/details?id=com.hpb.iquit1&hl=en_GB [accessed 2014-01-29]
61. iTunes. Rockville, MD: MMG, Inc QuitSTART URL: https://itunes.apple.com/kr/app/quitstart/id494552000?mt=8 [accessed 2014-01-29] [WebCite Cache ID 6MyvDJNJg]
62. Google Play. Brussels: Saatchi & Saatchi Exsmokers iCoach URL: https://play.google.com/store/apps/details?id=eu.stopsmokingcoach.icoach&hl=ko [accessed 2014-01-29]
63. Google Play. South Korea: Korean Ministry of Health and Welfare [No Smoke Guide] URL: https://play.google.com/store/apps/details?id=com.mgaon.nosmokeguide [accessed 2014-01-29] [WebCite Cache ID 6MywPIXEz]
64. Vansteenkiste M, Simons J, Lens W, Soenens B, Matos L, Lacante M. Less is sometimes more: goal content matters. Journal of Educational Psychology 2004;96(4):755-764. [doi: 10.1037/0022-0663.96.4.755]
65. Cornacchione J, Smith SW. The effects of message framing within the stages of change on smoking cessation intentions and behaviors. Health Commun 2012;27(6):612-622. [doi: 10.1080/10410236.2011.619252] [Medline: 22292861]

Abbreviations

SDT: self-determination theory

Edited by G Eysenbach; submitted 31.10.13; peer-reviewed by M Fiordelli, L Abroms, J Ferron, JM Montserrat; comments to author 21.11.13; revised version received 03.01.14; accepted 20.01.14; published 12.02.14.

Please cite as:
Choi J, Noh GY, Park DJ
Smoking Cessation Apps for Smartphones: Content Analysis With the Self-Determination Theory
J Med Internet Res 2014;16(2):e44
URL: http://www.jmir.org/2014/2/e44/
doi:10.2196/jmir.3061
PMID:24521881

Original Paper

Investigating the Congruence of Crowdsourced Information With Official Government Data: The Case of Pediatric Clinics

Minki Kim[1], PhD; Yuchul Jung[2], PhD; Dain Jung[1]; Cinyoung Hur[3], MS

[1]Department of Business and Technology Management, Korea Advanced Institute of Science and Technology, Daejeon, Korea, Republic Of
[2]Korea Institute of Science and Technology Information, Daejeon, Korea, Republic Of
[3]Electronics and Telecommunications Research Institute, Daejeon, Korea, Republic Of

Corresponding Author:
Minki Kim, PhD
Department of Business and Technology Management
Korea Advanced Institute of Science and Technology
KAIST N5-2109
291 Daehak-ro, Yuseong-gu
Daejeon, 305-701
Korea, Republic Of
Phone: 82 423506315
Fax: 82 423506339
Email: minki.kim@kaist.ac.kr

Abstract

Background: Health 2.0 is a benefit to society by helping patients acquire knowledge about health care by harnessing collective intelligence. However, any misleading information can directly affect patients’ choices of hospitals and drugs, and potentially exacerbate their health condition.

Objective: This study investigates the congruence between crowdsourced information and official government data in the health care domain and identifies the determinants of low congruence where it exists. In-line with infodemiology, we suggest measures to help the patients in the regions vulnerable to inaccurate health information.

Methods: We text-mined multiple online health communities in South Korea to construct the data for crowdsourced information on public health services (173,748 messages). Kendall tau and Spearman rank order correlation coefficients were used to compute the differences in 2 ranking systems of health care quality: actual government evaluations of 779 hospitals and mining results of geospecific online health communities. Then we estimated the effect of sociodemographic characteristics on the level of congruence by using an ordinary least squares regression.

Results: The regression results indicated that the standard deviation of married women’s education (*P*=.046), population density (*P*=.01), number of doctors per pediatric clinic (*P*=.048), and birthrate (*P*=.002) have a significant effect on the congruence of crowdsourced data (adjusted R^2=.33). Specifically, (1) the higher the birthrate in a given region, (2) the larger the variance in educational attainment, (3) the higher the population density, and (4) the greater the number of doctors per clinic, the more likely that crowdsourced information from online communities is congruent with official government data.

Conclusions: To investigate the cause of the spread of misleading health information in the online world, we adopted a unique approach by associating mining results on hospitals from geospecific online health communities with the sociodemographic characteristics of corresponding regions. We found that the congruence of crowdsourced information on health care services varied across regions and that these variations could be explained by geospecific demographic factors. This finding can be helpful to governments in reducing the potential risk of misleading online information and the accompanying safety issues.

(J Med Internet Res 2014;16(2):e29) doi:10.2196/jmir.3078

KEYWORDS

online health community; crowdsourcing; risk of misinformation; public health

Introduction

The advancement of information and communications technology has affected every aspect of modern society by enabling people to collectively create and exchange knowledge in pursuing their rights. In this era of Web 2.0, people can make smarter decisions with new interpretations and discoveries based on social interactions through social networking sites and online communities. With a basis in “trust your users,” Web 2.0 expedites knowledge acquisition through sharing and collaboration among Web users; that is, through harnessing collective intelligence [1]. Given this background, the questions that arise are (1) whether online users are trustworthy at all times, and (2) whether crowdsourced information benefits fields that require specialized knowledge, such as health care and medicine. In public health, it is important to investigate the accuracy of crowdsourced information within online health communities because any misleading information can directly affect patients’ choices of hospitals and drugs and can potentially exacerbate their health conditions. The social costs of inaccurate health information far exceed the financial costs incurred from poor choices of ordinary consumer products and services.

As the Web has become an important mass medium for consumers seeking health information and health care services online, researchers have focused on the advent of Health 2.0 or Medicine 2.0, which is the application of Web 2.0 technologies to health and medicine [2]. In the present era of Medicine 2.0, patients evaluate physicians and hospitals, share their health care experiences, and communicate with other patients through health-related websites [3-5].

Proliferation of these Medicine 2.0 sites (eg, PatientsLikeMe and Hello Health) benefits society by helping patients acquire knowledge on health through collaboration [3,4]. As a means of promoting health education, Medicine 2.0 has also reshaped doctor-patient relationships, making them more patient friendly [6].

To analyze such health information and communication patterns on websites and social media, considerable literature has accumulated over the years under the name of infodemiology [7,8], the science of distribution and determinants of information on the Internet aiming to inform public health and public policy.

In-line with infodemiology, researchers have attempted to understand and analyze unstructured free-text information available from the Internet, such as health-related user-generated data, or crowdsourced data. For instance, Marcus et al [9] searched Internet blogs that contained mental health-related keywords to analyze young adults’ concerns regarding mental health. More recently, Zhang et al [10] conducted sentiment analysis by using free-text comments about hospitals on websites and found that crowdsourced data are moderately associated with the paper-based national inpatient survey results in England.

On the other hand, concerning low quality information on the Internet that could harm public health [11], several studies have discussed the potential risk of inaccurate online information and accompanying safety issues. Using Google search results for severe acute respiratory syndrome (SARS) as examples, Morahan [12] argued that the overall quality of online health information is a serious concern for people because virtual access to high-quality health information is counterbalanced by high access to inaccurate and even fraudulent information. A systematic review on the quality of health information has found that 70% of studies on this subject have had concerns about the quality of crowdsourced health information. In fact, even too much information affects patients, making them even more confused because they find it harder to distinguish between true and inaccurate health information [13].

Hence, this study focused on the following: (1) whether crowdsourced information built in the online world is accurate, and (2) if imprecise, what drives this information inaccuracy. Similar to our study, Tsai et al [14] investigated social networking websites on consumer health and found that more than half of user-generated health information was incomplete or erroneous. Looking at drug information on Wikipedia, a different type of information source, Clauson et al [15] also found that there are more omission errors and less completeness in Wikipedia articles than Medscape Drug Reference articles. Consistent with these studies, we show the existence of inaccurate online information. Unlike previous research, however, this study adds to the literature by demonstrating heterogeneous inaccuracy between geospecific online health communities and objective hospital ratings, and by further delving into the offline determinants of such heterogeneity. To the best of our knowledge, no previous empirical study has taken this approach.

In this study, we investigate the quality of crowdsourced health information by evaluating conformity between 2 rankings: (1) government evaluations of hospitals, and (2) rankings obtained through mining of geospecific online health communities across South Korea

Methods

Overview

We employed 3 types of information source: (1) governmental evaluations of medical services aimed at ranking hospitals in the region, (2) rankings provided by online communities on hospitals according to the crowdsourced information of Web users, and (3) the census results by region provided by Statistics Korea to understand the influence of demographic features on the congruence between crowdsourced information and official government data in the health care domain.

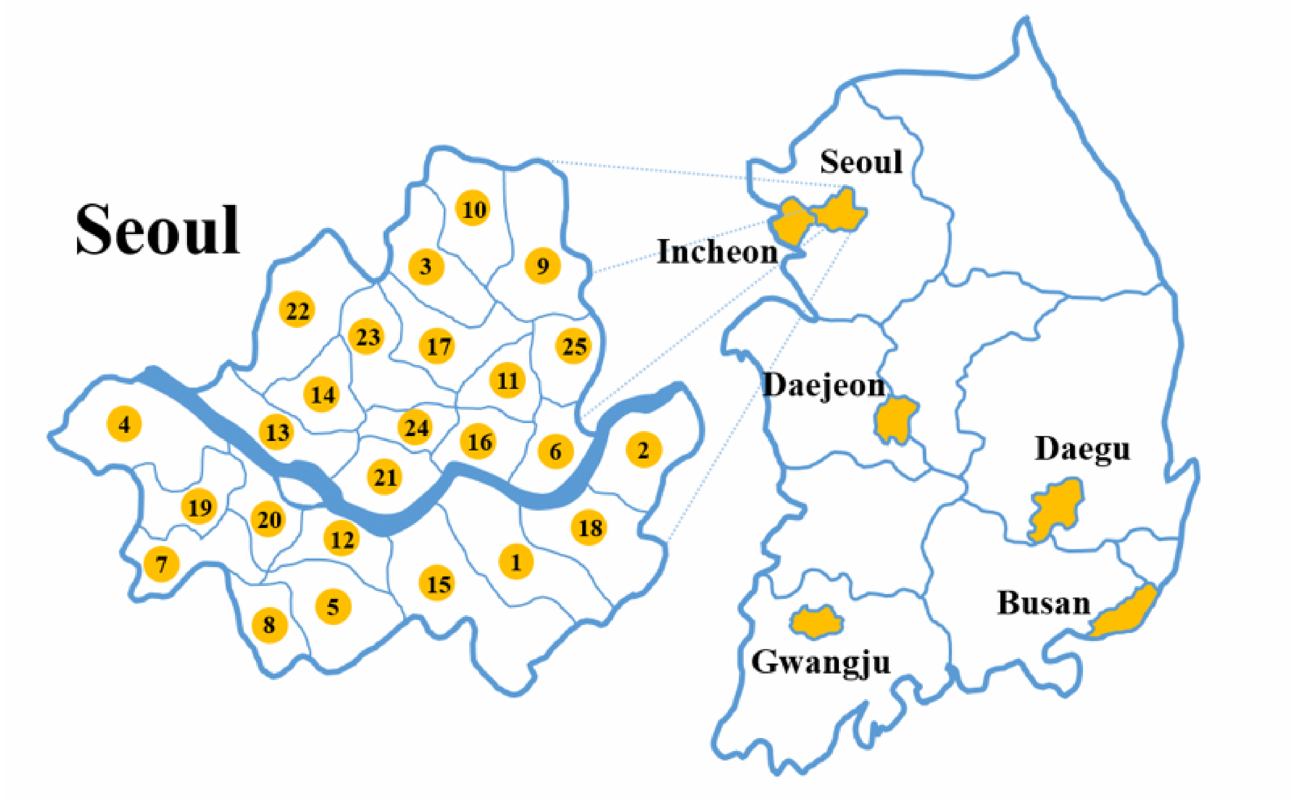
Government Hospital Evaluation Information

The sampled hospitals were confined to pediatric offices because young parents more eagerly seek information both online and offline, serving as agents for their children who can seldom self-diagnose the various symptoms they are suffering. According to Plantin and Daneback [16], first-time mothers aged 30-35 years are most active in searching for health and patient information on the Internet. Compared with other medical fields, parents of young children more actively share information on pediatrics over online health communities [17].

This study was conducted in the 6 major metropolitan cities of South Korea. Specifically, we examined 30 regions in total, including 25 districts (gu) within Seoul (accounting for one-quarter of the total population of South Korea), and 5 other metropolitan cities: Gwangju, Busan, Daegu, Daejeon, and Incheon. Figure 1 shows the geographic locations of the 30 regions.

As an objective measure of the medical service quality of local pediatricians, we relied on usage rates of antibiotics. The overuse of antibiotics is a global concern because it can result in severe adverse effects for children [18,19]. In fact, improper or excessive use of antibiotics also results in antibiotic-resistance problems, which, in turn, lead to unnecessary expenditures of public health care funding [20]. Korea is no exception in this issue. Antibiotic prescribing behavior has changed after quality assessment of prescriptions were done, including those for antibiotics for treatment of acute upper respiratory tract infections in ambulatory care in 2001 and the public reporting of its results in 2006 [21]. Since 2009, to discourage the excessive use of antibiotics, the Health Insurance Review and Assessment Service of Korea (HIRA) [22] has provided information on antibiotic usage rates for all hospitals across the nation to the public. The antibiotic usage rate is represented as the total number of antibiotic prescriptions over the total number of visits. We considered this usage rate as a hospital quality index, and based on these HIRA usage rates, we rank-ordered all 779 pediatric clinics in the sampled cities.

Figure 1. Geographic regions in South Korea associated with online health communities.



Mining Online Health Communities

To investigate the crowdsourced information for public health in South Korea, particularly on the quality of pediatric clinics, we focused on user-generated content from online communities hosted by Korean Web portals. Preliminary examination showed that online communities specializing in pediatrics served as places for parents to actively discuss (through questions and answers) the quality of pediatric clinics by sharing their thoughts, ideas, and experiences (ie, crowdsourcing). Naver [23] is a leading Internet search portal in Korea with an average market share of 78% and Daum [24] ranks second with an average market share of 15.22% in 2013 [25], and together they maintain the largest online health communities for child-raising parents; therefore, we analyzed the contents of only the local online health communities hosted by Naver and Daum. Furthermore, these 2 sites maintain various online health communities by topic and by region across the nation, such as support groups for cancer patients and general parental care in each local region.

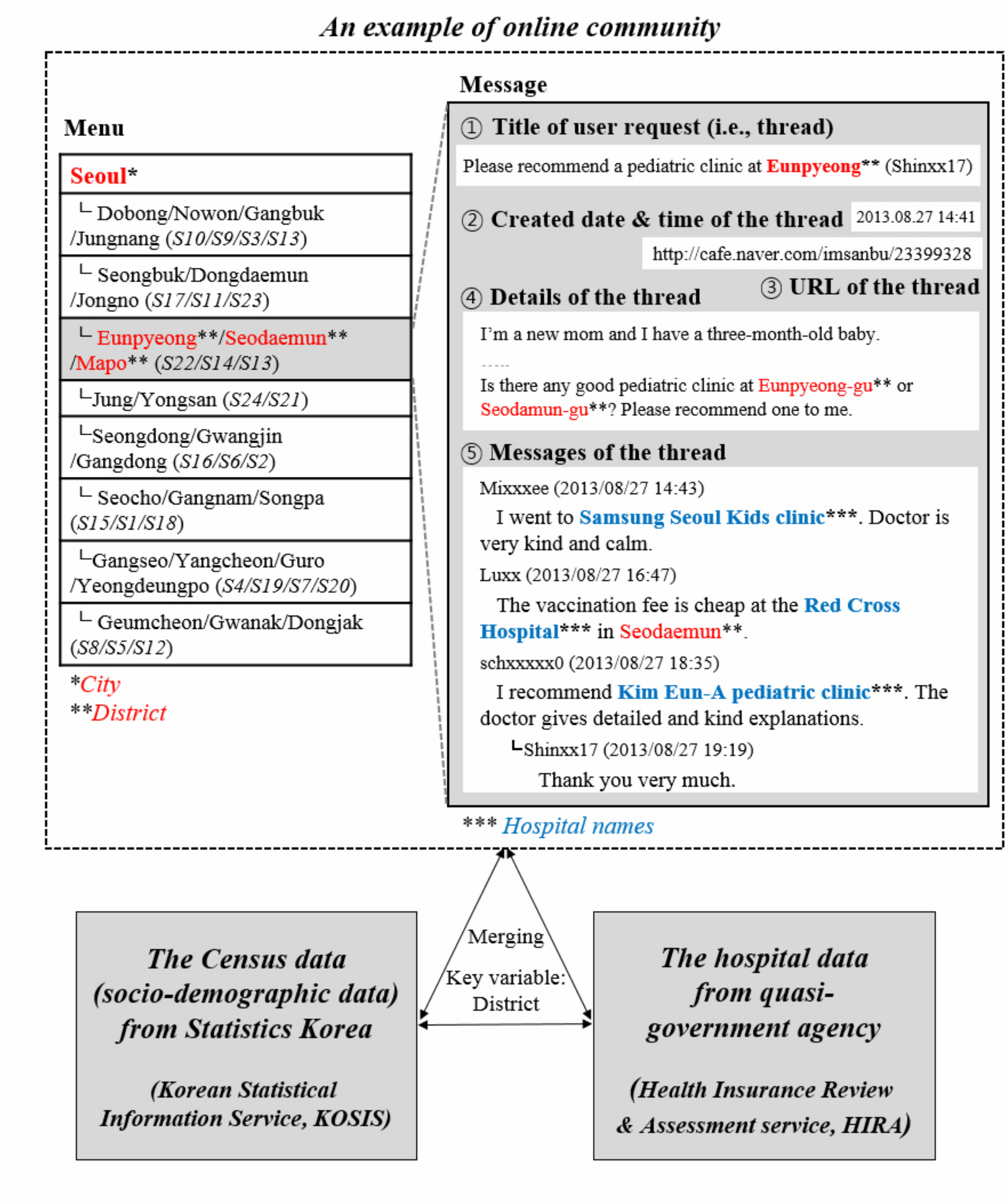
In particular, we were interested in the frequencies at which hospital names appeared in discussions on these websites. To ensure that frequently mentioned hospitals were more likely to be recognized as providing high-quality care, we only considered users' responses to others' requests for hospital recommendations. A certain user response could denote positive or negative attitudes based on his/her experiences. Sometimes, however, these attitudes were neutral (ie, only hospital names were mentioned without sentimental attitudes) or ambiguous (ie, positive and negative attitudes coexisted). Based on our observation of approximately 18.45% of the sample (32,065/173,748), 92.03% (29,511/32,065) were identified as being positive or neutral in attitude. In this context, we regarded the frequency of hospital names appearing in positive and neutral mentions as a measure of crowdsourced information, representing parents' beliefs about the quality of local hospitals. In this study, we did not consider negative mentions about hospitals in measuring crowdsourced information because their occurrence rate was less than 7.97% (2554/32,065) and including them did not affect the results of hospital evaluations. To aggregate the user-generated content that might contribute to constructing geospecific crowdsourced information on this particular topic, we selected candidate online communities on the basis of the number of members and the numbers of threads and messages. The contents of online communities can be divided into 3 levels: threads, messages, and sentences. Threads often contain several messages, whereas messages are short, often composed of only a few sentences or sentence fragments.

Figure 2 is a snapshot of the hospital recommendation contents in an online community.

Table 1 shows the statistics of the content of the selected online communities for parents on Naver and Daum. We obtained 32,422 threads and over 170,000 messages. Thus, there were an average of 5.36 pediatric hospital–related discussions per thread among all community members.

To mine discussions on pediatric hospitals from the 2 major Web portals, we developed a text-mining tool tailored to handle the difficulties inherent to the Korean language and the characteristics of local online communities for parents, as shown in Figure 3.

Figure 2. An example of an online community and the data construction process.



Please recommend a pediatric clinic at Eunpyeong** (Shinxx17)

2013.08.27 14:41

http://cafe.naver.com/imsanbu/23399328

I'm a new mom and I have a three-month-old baby.

.....

Is there any good pediatric clinic at Eunpyeong-gu** or Seodamun-gu**? Please recommend one to me.

Mixxxee (2013/08/27 14:43)
I went to **Samsung Seoul Kids clinic*****. Doctor is very kind and calm.

Luxx (2013/08/27 16:47)
The vaccination fee is cheap at the **Red Cross Hospital***** in Seodaemun**.

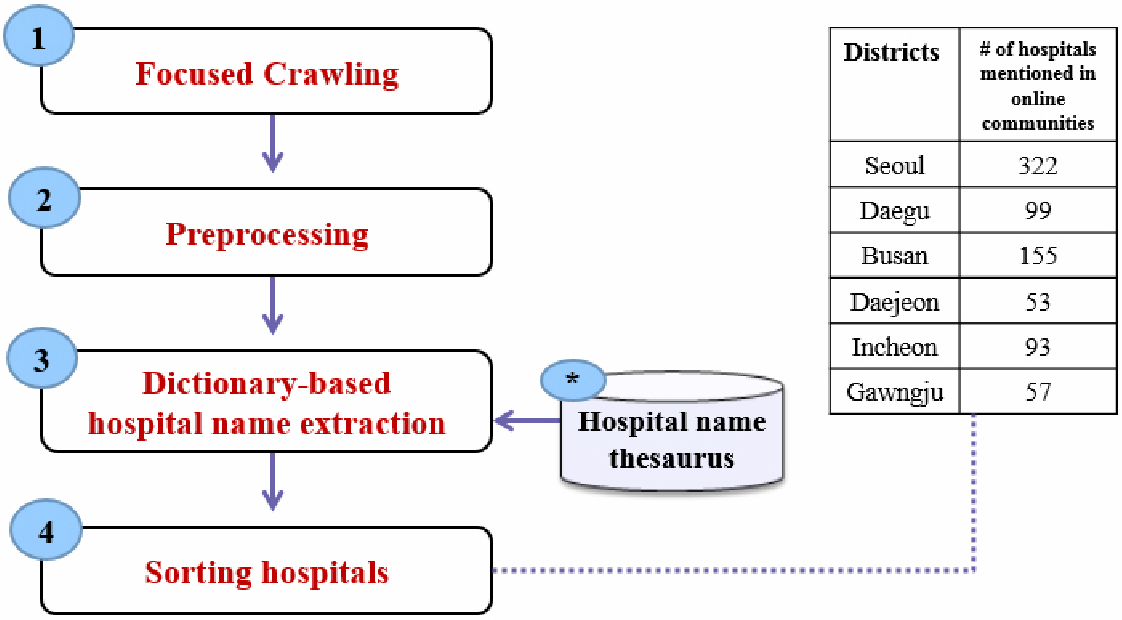
schxxxxx0 (2013/08/27 18:35)
I recommend **Kim Eun-A pediatric clinic*****. The doctor gives detailed and kind explanations.

└Shinxx17 (2013/08/27 19:19)
Thank you very much.

Table 1. Characteristics of the data extracted from selected online communities hosted by Naver and Daum Web portals.

District	Threads, n	Messages, n	Messages per thread, mean
Seoul	10,832	54,392	5.02
Daegu	8072	47,419	5.87
Busan	5965	28,910	4.85
Daejeon	3952	22,475	5.69
Incheon	775	5184	6.69
Gwangju	2826	15,368	5.44
Total	32,422	173,748	5.36

Figure 3. Text mining for hospital name extraction.



Districts	# of hospitals mentioned in online communities
Seoul	322
Daegu	99
Busan	155
Daejeon	53
Incheon	93
Gawngju	57

Step 1: Focused Crawling

The previously mentioned Web portals, Naver and Daum, operate various kinds of online communities by region and subject. For example, the Naver portal has approximately 2 million communities by region. We observed that each region included at least 1 parents' community with outstanding members, popularity, and exhibited vigorous use by the members rather than others. We selected these local online communities as our text-mining information sources. Each online community provided a search function that helped us identify Web pages with the keywords "pediatric" and "recommendation." The Web pages from selected online communities of parents were crawled and stored in hypertext markup language (HTML) format.

Step 2: Preprocessing

We deleted HTML tags and nontextual information, such as images, JavaScript codes, and advertisements from the extracted files. For effective handling of HTML content, we used the Beautiful Soup library, designed in the Python programming language [26]. In addition, we only selected candidate messages that recommended or in some way mentioned hospitals to others.

Step 3: Dictionary-Based Hospital Name Extraction

When people mention hospitals in a social context in Korean, they can use a range of hospital names, mostly expressed through acronyms or abbreviations. To counter this problem, we built a dictionary of hospital names covering the pediatric hospitals in the 6 South Korean metropolitan cities (ie, Seoul, Daegu, Busan, Daejeon, Gwangju, and Incheon) based on the HIRA website. In the hospital name thesaurus, each hospital name had 3 similar expressions on average. This dictionary was a value (expression) mapping table that added acronyms or abbreviations of every hospital name observed in user-written sentences in the online communities. Using this dictionary, our text-mining module performed a specially designed stepwise expression normalization procedure based on textual similarity computed by edit-distance [27] between the synonyms in the dictionary and the expressions in the sentences. This was used to find canonical hospital names for use in raw hospital representations.

Step 4: Sorting Hospitals

After we extracted the canonical hospital names from messages, we calculated the hospital name frequency by district. After sorting the names by descending order, we could compare the ranked hospital information with the hospital quality rankings obtained from the HIRA, which were based on antibiotic usage rate.

We assessed the accuracy of the hospital name extraction through human tagging of a random sample of 9450 messages (5.44%) of the total messages (n=173,748). We achieved a recall (the proportion of hospital names in the original text that were extracted correctly) of 73% and precision (the proportion of hospital names extracted that were correct) of 84%. We calculated the F_1 score, the harmonic mean of precision and recall, as F_1=2×(precision)×(recall)/(precision+recall)=77.7%. Although accuracy measures of 80% to 90% are often achieved in named entity extraction in English, such high accuracy in informal texts written in Korean is rare because of various types of acronyms, numbers of misspellings, and frequently appearing incorrect spacing between words. Together with the fact that Korean is an agglutinative language, these factors act as bottlenecks with reference to performance.

Measuring the Congruence of Crowdsourced Data

To measure the congruence of crowdsourced data with official government data, we first compared hospital ratings based on antibiotics prescriptions with hospital rankings based on crowdsourced information from online health communities. The congruence of crowdsourced data is then determined based on the conformity of both rankings. For this, we used the following 2 indexes: Kendall tau [28] and Spearman rho rank correlation coefficients. Kendall tau is defined as Equation (1) and Spearman rho is calculated as Pearson correltation coefficient based on ranks and average ranks as Equation (2) in Figure 4.

Both Kendall tau and Spearman rho range from –1 to +1. A value of +1 means that the 2 rankings completely agree and a value of –1 means that the 2 rankings are completely contradictory. Thus, the larger the index value, the greater the correspondence between the objective hospital ranking and crowdsourced ranking, which implies that online users are developing correct crowdsourced information.

Figure 4. Definitions of Kendall tau and Spearman rho rank correlation coefficients.

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \qquad (1)$$

where n represents the sample size,
n_c is the number of concordant pairs,
n_d is the number of discordant pairs.

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad (2)$$

where d_i is the difference in rankings for each object i, $i \in \{1, 2, \ldots, n\}$.

Geospecific Sociodemographic Information

Considering that information exchanges within online health communities occur disparately and are bounded by geographic region, we can examine the demographic information in a given region to identify the particular sociodemographics that influence the congruence of crowdsourced data (ie, the degree of correspondence between government ratings of hospitals in the region and Internet rankings). We selected 6 variables for use in the analysis: birthrate by region, mean and standard deviation of educational attainment for married women, population density, number of pediatric clinics per geographic area, and mean number of doctors per pediatric office. On the basis of previous studies that found pregnant and first-time mothers actively search for online health information, we selected birthrate as the independent variable [16,29]. Gender and education have been found to affect online health searching, specifically women and more educated individuals were found to conduct more online searches [17,30-32]. Therefore, education of married women was selected as a variable. In addition, we chose 3 other variables for availability of pediatric clinics and possible variation of assessment of doctors. Educational attainment refers to the highest level of education that married women had completed and was categorized into no education, elementary school, middle school, high school, 2-year college, 4-year college, master's degree, and PhD. These were then converted into values of 0, 6, 9, 12, 14, 16, 18, and 23 years of schooling, respectively, to calculate the average and standard deviations for education. For population density, we divided the population of each region by its area. Finally, for the number of doctors per pediatric clinic, we divided the total number of pediatric doctors in the region by the total number of pediatric clinics. We also used another variable, availability, to describe the availability of medical services, which was computed by dividing the total number of pediatric clinics in the region by the total population of that region. Finally, birthrate refers to the average number of children per fertile woman. Tables 2 and 3 describe the variables along with appropriate summary statistics and correlations between independent variables.

For hospitals recommended within local online communities, grouping them by year could have led to difficulties in comparing rankings because of the limited number of observations per local area. Therefore, we aggregated the number of recommended hospitals in a given region between 2009 and 2012 to compile the rankings. For comparison, we also calculated the average rate of antibiotic prescriptions between 2009 and 2012 to determine the objective hospital ranking.

For the population, area (km^2) of a region, and birthrate data, we computed the means for the data provided in the Korean Statistical Information Service (KOSIS) [33] for a given period and used them as sociodemographic variables. However, for educational attainment, we had access to 2010 data only, and thus used them as they were. As the number of hospitals and the number of doctors per clinic in each year were unknown, we used the information registered at the HIRA website for 2012. We present all the sources of data in Table 4.

To analyze which sociodemographic features influenced the congruence of crowdsourced data, we ran an ordinary least squares (OLS) regression on the 29 regions, with rank correlation as the dependent variable and the aforementioned 6 sociodemographic characteristics as independent variables.

Table 2. Summary statistics of variables.

Variable	Description	Mean	SD	Min	Max
Birthrate	Birthrate (number of births per fertile woman)	1.048	0.100	0.855	1.261
Education mean	Average educational attainment of married women	12.390	0.833	11.340	14.858
Education SD	Standard deviation of educational attainment of married women	3.496	0.183	3.149	3.788
Population density	Population per area (km^2)	0.016	0.007	0.003	0.029
Doctors per clinic	Number of doctors per pediatric clinic	1.641	0.396	1.167	2.889
Availability	Number of pediatric clinics per 1000 persons	0.060	0.010	0.033	0.080

Table 3. Correlations between independent variables.

Pearson correlation	Birthrate		Education mean		Education SD		Population density		Doctors per hospital		Availability	
	r	*P*	*r*	*P*	*r*	*P*	*r*	*P*	*r*	*P*	*r*	*P*
Birthrate	1.00											
Education mean	–.27	.16	1.00									
Education SD	.03	.88	–.34	.07	1.00							
Population density	–.58	.001	.03	.89	–.10	.60	1.00					
Doctors per hospital	–.23	.24	.27	.15	.22	.24	–.06	.74	1.00			
Availability	.11	.57	.38	.04	.09	.66	–.15	.44	.26	.17	1.00	

Table 4. Data sources and time periods used.

Information type and source	Data	Period
Official government data		
Health Insurance Review & Assessment Service (HIRA)	Antibiotic prescription rate	2009-2012 (Average)
	# of hospitals	2013
	# of doctors	2013
Crowdsourced data from online communities		
Web portals (Daum, Naver)	# of recommended hospitals	2009-2012 (Sum)
Geospecific demographic information		
Korean Statistical Information Service (KOSIS)	Population	2009-2012 (Average)
	Area of a region	2009-2012 (Average)
	Birthrate	2009-2012 (Average)
	Education	2010

Results

Rank Correlations

Table 5 shows the rank correlation coefficients by region for 29 of the 30 regions. We excluded Jongno-gu in Seoul because it had relatively few online communities.

As seen in Table 5, significant variations in Kendall tau and Spearman rho scores exist across regions, with the lowest for region S7 and highest for region S8.

Table 5. Congruence of geospecific crowdsourced health information.

District[a]	n[b]	Kendall tau	Rank	Spearman rho	Rank
S1	14	0.132	11	0.152	11
S2	21	0.214	8	0.292	9
S3	11	0.255	7	0.330	7
S4	15	0.038	18	0.040	18
S5	15	0.438	2	0.587	2
S6	17	0.265	6	0.415	6
S7	11	–0.346	28	–0.466	28
S8	5	0.5	1	0.667	1
S9	22	0.052	15	0.093	12
S10	10	–0.089	23	–0.109	22
S11	11	0.382	3	0.441	5
S12	12	–0.182	25	–0.232	24
S13	15	0.210	9	0.321	8
S14	13	–0.077	22	–0.135	23
S15	11	0.327	5	0.471	3
S16	11	–0.236	27	–0.340	27
S17	10	–0.156	24	–0.268	25
S18	26	0.04	16	0.055	17
S19	18	–0.235	26	–0.326	26
S20	15	0.086	12	0.080	13
S21	8	–0.357	29	–0.470	29
S22	16	0.35	4	0.458	4
S24	3	0.00	20	0.000	20
S25	12	0.061	13	0.078	14
Gwangju	57	0.040	16	0.060	16
Daegu	99	0.180	10	0.268	10
Daejeon	53	0.025	19	0.040	18
Busan	155	0.056	14	0.078	14
Incheon	93	–0.006	21	–0.001	21
Total	779				

[a] S means Seoul and S1 represents district (gu) 1 of Seoul. S23 is Jongno-gu. Matched district names are listed in Figure 2.

[b] Number of pediatric clinics that were mentioned on online communities.

Ordinary Least Squares Regression Results

Table 6 shows the OLS regression results that we obtained (adjusted R^2=.334). All independent variables, except availability, had negative coefficient values. Birthrate and number of doctors per clinic significantly lowered the congruence of crowdsourced data with official government data at the 1% level, whereas the standard deviation of educational attainment and population density negatively affected congruence at a significance level of 5%. However, the variables of average education attainment and number of hospitals per 1000 people were not statistically significant. In terms of each independent variable, our study showed that the higher the birthrate in a given region, the larger the standard deviation of educational attainment, the higher the population density, and the greater the number of doctors per clinic, the more likely that the crowdsourced data within online communities in a given region was less congruent. We observed similar findings when we used the Spearman rho rank correlation coefficient instead of Kendall tau for ranking comparisons in the OLS regression analysis (adjusted R^2=.364).

Table 6. Parameter estimates and standard errors.

Independent variable	Kendall tau			Standardized coefficient[a]	Spearman rho		
	Coefficient	SE	*P* value		Coefficient	SE	*P* value
Birthrate	–1.678	0.477	.002	–0.748	–2.302	0.633	.001
Education mean	–0.091	0.055	.12	–0.335	–0.117	0.073	.12
Education SD	–0.471	0.223	.046	–0.382	–0.636	0.296	.04
Population density	–16.860	6.244	.01	–0.532	–24.077	8.282	.008
Doctors per clinic	–0.212	0.101	.048	–0.373	–0.311	0.134	.03
Availability	2.611	4.077	.53	0.116	3.797	5.407	.49
Constant	5.051	1.407	.002		6.837	1.866	.001

[a] Standardized coefficients are the regression coefficients obtained by first standardizing all variables to have a mean of 0 and a standard deviation of 1.

We also standardized the coefficients to better compare the impact of the variables. Birthrate had the largest standardized coefficient whereas doctors per clinic had the smallest. We found that we could interpret the value of standardized coefficients (ie, a 1 SD decrease in birthrate led to a 0.75 SD increase in predicted Kendall tau, with the other variables held constant). Similarly, a 1 SD decrease in doctors per clinic yielded a 0.37 SD increase in the predicted Kendall tau. These results, presented in Table 6, can explain the variations in Kendall tau values across districts we observed in Table 5.

For illustration purposes, we compared 2 districts in Seoul: S5 (with a relatively high Kendall tau of 0.44) and S18 (with a low Kendall tau of 0.04). Large differences in demographic characteristics between S5 and S8 in birthrate (0.897 vs 1.042) and doctors per clinic (1.185 vs 1.808) explained why S5 had a higher Kendall tau value. Similarly, we could explain the difference in congruence of crowdsourced data between the local cities. For instance, a higher Kendall tau value for Daegu in comparison to that for Daejeon was associated with Daegu's lower birth rate, standard deviation of educational attainment, and number of doctors per clinic.

A higher birthrate in a region implies that there is greater percentage of women with children under age 2 years who have relatively less experience with pediatric clinics. As a result, the proportion of women seeking and sharing information on pediatric hospitals in online communities may be greater for regions with higher birthrates compared with other regions. This can explain why the congruence of crowdsourced data is likely to be lower in a region with higher birthrates. Interestingly, our findings also showed that the average educational attainment of married women in a region was not significantly related to the level of congruence; however, this was not the case for the standard deviation of educational attainment, which had a significant negative relationship with the congruence of crowdsourced data. This difference may be because of the potential mingling of correct and misleading information in the exchange of hospital information in online communities, where users do not have any knowledge of the educational attainment of other users. It is also likely that more densely populated regions are more likely to have a greater number of people gathering pediatric hospital information through offline channels (ie, nearby friends or relatives) than through online communities. In other words, crowdsourced information is created from face-to-face interactions—an offline channel—meaning that there is no definitive build-up of online crowdsourcing in this case. Finally, the greater the number of doctors per clinic, the higher the chances of people may build up less congruent crowdsourced information on these hospitals. This can be explained by the notion that people often judge the same hospital differently according to the doctor they visit; thus, there would be a greater variations in users' evaluations of a hospital with a large number of doctors as opposed to a smaller number.

Discussion

Although patients actively create and exchange health information in this era of Health 2.0, the spread of inaccurate health information can have a potentially negative effect on public health because of its direct impact on patients' health conditions. Hence, in this study, we investigated the congruence between crowdsourced information and official government data in the health care domain and geospecific determinants of misinformation across South Korea. In particular, this is the first attempt in the literature to associate the results of data mining from geospecific online health communities with sociodemographic characteristics of regions.

We found that the quality of crowdsourced information on health care services varies across regions and that these variations can be explained by geospecific demographic factors, such as birthrate, educational attainment, population density, and the number of doctors per clinic. The findings of our study hold practical implications for health policies. From our research, we suggest that government authorities should recognize the power of crowdsourcing and make efforts to reduce the potential risk of low quality health information and the accompanying safety issues. Crowdsourcing for health care services, unlike that for general products and services, is based on contributions from nonspecialists. Furthermore, the quality of crowdsourced information can be lower for pediatric clinics in particular because parents serve as agents for their children who can seldom self-diagnose the various symptoms they are suffering. Therefore, our findings also suggest the need for governmental efforts to counterbalance the misinformation by disseminating approval ratings of pediatric clinics via geospecific online health communities, and that this can be accomplished by prioritizing the regions with the lowest information quality levels.

This study also had some limitations. Although we collected data on major online communities with many active users by region, we did not fully evaluate all hospitals in each region. However, as we confined our focus to online communities on pediatrics with a relatively large number of users—consisting of typical mothers of young children—we can say that the hospital list by region extracted using text-mining techniques was quite reliable in practice. In addition, although the text mining showed reasonable performance in the Korean language, performance could be enhanced through anaphora resolution of hospital names. It should be noted that the context of an Internet forum is disorganized and the language used tends to follow the idiosyncratic nature of human beings rather than grammatical standards and rules. Therefore, in our future research, we plan to extend the application of the text-mining tool to sentiment analysis and the specific textual relationships between qualitative factors that influence people's choice of hospitals. In that way, we hope we can draw more practical implications for health policies.

Conflicts of Interest

None declared.

References

1. 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 2014-01-19] [WebCite Cache ID 6MiXNbMo1]
2. Eysenbach G. Consumer health informatics. BMJ 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [Medline: 10864552]
3. 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]
4. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009 Mar;28(2):361-368 [FREE Full text] [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
5. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
6. Popoiu MC, Grosseck G, Holotescu C. What do we know about the use of social media in medical education? In: Procedia - Social and Behavioral Sciences. 2012 Jan Presented at: 4th World Conference on Educational Sciences (WCES-2012) 02-05; February 2012; Barcelona, Spain p. 2262-2266. [doi: 10.1016/j.sbspro.2012.05.466]
7. Eysenbach G. Infodemiology: the epidemiology of (mis)information. The American Journal of Medicine 2002 Dec 15;113(9):763-765. [doi: 10.1016/S0002-9343(02)01473-0] [Medline: 12517369]
8. 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 Mar;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
9. Marcus MA, Westra HA, Eastwood JD, Barnes KL, Mobilizing Minds Research Group. What are young adults saying about mental health? An analysis of Internet blogs. J Med Internet Res 2012 Jan;14(1):e17 [FREE Full text] [doi: 10.2196/jmir.1868] [Medline: 22569642]
10. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
11. Eysenbach G, Powell J, Kuss O, Sa E. Empirical Studies Assessing the Quality of Health Information for Consumers on the World Wide Web. JAMA 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
12. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [Medline: 15667044]
13. 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]

14. Tsai CC, Tsai SH, Zeng-Treitler Q, Liang BA. Patient-centered consumer health social network websites: a pilot study of quality of user-generated health information. In: AMIA Annu Symp Proc. 2007 Oct Presented at: Proceedings of the Annual Symposium of the American Medical Informatics Association; 2007; Chicago, IL p. 1137.
15. Clauson KA, Polen HH, Boulos MN, Dzenowagis JH. Scope, completeness, and accuracy of drug information in Wikipedia. Ann Pharmacother 2008 Dec;42(12):1814-1821. [doi: 10.1345/aph.1L474] [Medline: 19017825]
16. Plantin L, Daneback K. Parenthood, information and support on the internet. A literature review of research on parents and professionals online. BMC Fam Pract 2009;10:34 [FREE Full text] [doi: 10.1186/1471-2296-10-34] [Medline: 19450251]
17. Stern MJ, Cotten SR, Drentea P. The separate spheres of online health: gender, parenting, and online health information searching in the information age. Journal of Family Issues 2011 Oct 21;33(10):1324-1350. [doi: 10.1177/0192513X11425459]
18. Weissman J, Besser RE. Promoting appropriate antibiotic use for pediatric patients: a social ecological framework. Semin Pediatr Infect Dis 2004 Jan;15(1):41-51. [Medline: 15175994]
19. Paluck E, Katzenstein D, Frankish CJ, Herbert CP, Milner R, Speert D, et al. Prescribing practices and attitudes toward giving children antibiotics. Can Fam Physician 2001 Mar;47:521-527 [FREE Full text] [Medline: 11281085]
20. Mainous AG, Hueston WJ, Love MM, Evans ME, Finger R. An evaluation of statewide strategies to reduce antibiotic overuse. Fam Med 2000 Jan;32(1):22-29. [Medline: 10645510]
21. Kim B. Overview of antibiotic use in Korea. Infect Chemother 2012;44(4):250. [doi: 10.3947/ic.2012.44.4.250]
22. Health insurance review and assessment service. URL: http://www.hira.or.kr/ [accessed 2013-10-28] [WebCite Cache ID 6Ki5geHrV]
23. NAVER Corp. URL: http://www.naver.com/ [accessed 2013-10-28] [WebCite Cache ID 6Ki54BAEM]
24. Daum communications. URL: http://www.daum.net/ [accessed 2013-10-28] [WebCite Cache ID 6Ki7w7xVe]
25. Internet Trend. Search engine market share URL: http://www.internettrend.co.kr/trendForward.tsp [accessed 2013-10-28] [WebCite Cache ID 6Ki6hA4Km]
26. Leonard R. Beautiful Soup. URL: http://www.crummy.com/software/BeautifulSoup/ [accessed 2013-10-28] [WebCite Cache ID 6Ki6tNAXD]
27. Vilares M, Ribadas FJ, Vilares J. Phrase similarity through the edit distance. In: Database and Expert Systems Applications. Berlin: Springer; 2004:306-317.
28. Kendall MG. A new measure of rank correlation. Biometrika 1938 Jun;30(1/2):81-93.
29. Daneback K, Plantin L. Research on parenthood and the internet: themes and trends. Cyberpsychology: Journal of Psychosocial Research on Cyberspace 2008;2(2):article2 [FREE Full text]
30. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
31. 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]
32. Hale TM, Cotten SR, Drentea P, Goldner M. Rural-urban differences in general and health-related internet use. American Behavioral Scientist 2010 Mar 09;53(9):1304-1325. [doi: 10.1177/0002764210361685]
33. Statistics Korea. Korean statistical information service URL: http://kosis.kr/ [accessed 2013-10-28] [WebCite Cache ID 6Ki7kxFCu]

Abbreviations

HIRA: Health Insurance Review and Assessment
HTML: hypertext markup language
KOSIS: Korean Statistical Information Service
OLS: ordinary least squares
SARS: severe acute respiratory syndrome

Edited by G Eysenbach; submitted 03.11.13; peer-reviewed by J Choi, H Tak, J Lee, S Erdley; comments to author 23.12.13; revised version received 07.01.14; accepted 09.01.14; published 03.02.14.

Please cite as:
Kim M, Jung Y, Jung D, Hur C
Investigating the Congruence of Crowdsourced Information With Official Government Data: The Case of Pediatric Clinics
J Med Internet Res 2014;16(2):e29
URL: http://www.jmir.org/2014/2/e29/
doi:10.2196/jmir.3078
PMID:24496094

Original Paper

The Use and Significance of a Research Networking System

Maninder Kahlon[1], PhD; Leslie Yuan[1], MPH; John Daigre[1]; Eric Meeks[1]; Katie Nelson[1], MBA; Cynthia Piontkowski[1]; Katja Reuter[2], PhD; Rachael Sak[1], RN, MPH; Brian Turner[1], MBA; Griffin M Weber[3], MD, PhD; Anirvan Chatterjee[1]

[1]University of California, San Francisco, Clinical & Translational Science Institute, San Francisco, CA, United States

[2]University of Southern California, Southern California Clinical & Translational Science Institute, Los Angeles, CA, United States

[3]Harvard Medical School, Harvard University, Boston, MA, United States

Corresponding Author:
Maninder Kahlon, PhD
University of California, San Francisco
Clinical & Translational Science Institute
185 Berry St, Suite 5300
San Francisco, CA, 94107
United States
Phone: 1 415 514 8060
Fax: 1 415 514 8520
Email: mini.kahlon@ucsf.edu

Abstract

Background: Universities have begun deploying public Internet systems that allow for easy search of their experts, expertise, and intellectual networks. Deployed first in biomedical schools but now being implemented more broadly, the initial motivator of these research networking systems was to enable easier identification of collaborators and enable the development of teams for research.

Objective: The intent of the study was to provide the first description of the usage of an institutional research “social networking” system or research networking system (RNS).

Methods: Number of visits, visitor location and type, referral source, depth of visit, search terms, and click paths were derived from 2.5 years of Web analytics data. Feedback from a pop-up survey presented to users over 15 months was summarized.

Results: RNSs automatically generate and display profiles and networks of researchers. Within 2.5 years, the RNS at the University of California, San Francisco (UCSF) achieved one-seventh of the monthly visit rate of the main longstanding university website, with an increasing trend. Visitors came from diverse locations beyond the institution. Close to 75% (74.78%, 208,304/278,570) came via a public search engine and 84.0% (210 out of a sample of 250) of these queried an individual’s name that took them directly to the relevant profile page. In addition, 20.90% (214 of 1024) visits went beyond the page related to a person of interest to explore related researchers and topics through the novel and networked information provided by the tool. At the end of the period analyzed, more than 2000 visits per month traversed 5 or more links into related people and topics. One-third of visits came from returning visitors who were significantly more likely to continue to explore networked people and topics (*P*<.001). Responses to an online survey suggest a broad range of benefits of using the RNS in supporting the research and clinical mission.

Conclusions: Returning visitors in an ever-increasing pool of visitors to an RNS are among those that display behavior consistent with using the tool to identify new collaborators or research topics. Through direct user feedback we know that some visits do result in research-enhancing outcomes, although we cannot address the scale of impact. With the rapid pace of acquiring visitors searching for individual names, the RNS is evolving into a new kind of gateway for the university.

(J Med Internet Res 2014;16(2):e46) doi:10.2196/jmir.3137

KEYWORDS

social networking; search engine; information seeking behavior; interprofessional relations; multidisciplinary communication

Introduction

New tools are enabling the search and discovery of researchers, their expertise, intellectual output, and professional networks. These Web-based applications mine a variety of data sources to automatically generate searchable profiles and expose existing networks of collaborators. In addition to large-scale commercial services [1], universities have begun to deploy such systems locally. Medical schools were early adopters, motivated by the promise of these tools in enabling the development of diverse research teams to meet the evolving emphasis of funders and the demands of translational science [2,3]. Momentum for these research "social networking" systems or research networking systems (RNSs) has continued to build; a recent pilot of a federated search of expertise counted 57 institutional participants [4,5]. Although they originated in biomedicine, the systems are now being extended to represent diverse institutional portfolios of research expertise [6,7].

RNSs deployed at medical schools can be contextualized against industry deployments of "expertise location" and "social networking" systems [8]. Like expertise location systems, RNSs enable the discovery of individuals based on their expertise using the automated generation of rich searchable profiles. Like enterprise social networking systems, RNSs allow for the search and browsing of networks of people and topics. Such systems are being deployed internally by large intellectual capital-driven companies (such as IBM [9] or Deloitte Consulting [10]) for reasons similar to those motivating the deployment of RNSs in academic settings (ie, to facilitate a better understanding of who's doing what and to enable knowledge sharing and team building [11]). But, RNSs are different in one significant way. Because they support the academic mission and because research increasingly requires collaboration beyond institutional boundaries [12], they are deployed as primarily public systems. As a result, unlike closed enterprise systems, a public RNS exposes rich content about the people in an institution to the broadest possible audience.

With an increasing number of RNSs deployed in academic biomedicine, there is now a focus on understanding how these tools are used. In industry, studies show that such tools are used predominantly to search for people [13] including those outside known circles of colleagues [9,14]. This is in contrast to findings from studies of users of Facebook where evidence suggests that users primarily reinforce existing networks, though this has only been studied in student populations [15,16]. In the academic setting, only one study reports use of a small private RNS showing that visitors who do log on spend more time in their session than comparable benchmarks for time spent on Google sessions [17]. In addition, several studies have documented needs and requirements for such systems to enable collaboration in science [18-20]. Outside of an organizational setting, LinkedIn's membership grew to more than 200 million members in 2013 [21], providing general testament to the perceived value of professional networking tools.

In 2010, the University of California, San Francisco (UCSF) became the first external adopter of an open source RNS called "Profiles Research Networking Software" [22]. After several years of deployment, we present a description of the usage of this publicly accessible RNS that exposes rich content about biomedical researchers at the institution to the broadest possible audience. Our objective is to provide the first description of the degree of usage of a public institutional research networking system and to identify sources of visitors, rate of engagement (visitors returning), degree of engagement (depth of visit), and outcomes as reported by users. We also wanted to assess evidence of usage of the more unique aspects of the RNS that allow users to view related people and concepts, exposing them to the connections that might induce identification of new collaborators or research themes. And finally, we were curious about how the introduction of a public social networking system might transform access to a university's primary intellectual capital—its people.

Methods

Ethics: Human Subjects Research

The study plan was submitted to the UCSF Institutional Review Board (Human Research Protection Program) and was determined not to be human subjects research.

UCSF Profiles Research Networking System

A screenshot of the UCSF Profiles RNS is shown in Figure 1. Multiple sources of data populate the contents of individual profile pages. These sources include publication feeds from the National Library of Medicine's PubMed, disambiguated to match author names to institutional sources of information [22]. UCSF Profiles extends the open source Profiles RNS in multiple ways, including integrating relevant commercial services such as YouTube and SlideShare.

Figure 1. Screenshot of profile in UCSF Profiles. Subareas A, B, C denote 3 sets of links that when clicked through count as an additional page on site. Subareas A1 and A2 denote links connecting visitors to related people and topics.

Sign in to edit your profile (add interests, mentoring, photo, etc.)

Jane Smith, PhD

Title Associate Professor in Residence
School UCSF School of Medicine
Department Medicine

Awards and Honors | Overview | Interests | Websites | Featured Videos | In The News
Faculty Mentoring | Awarded Grants | Global Health | Twitter | More Info | Publications

Awards and Honors
2012 Lee Goldman, MD Endowed Chair in Medicine
2010 American Society for Clinical Investigation

Overview
Jane Smith, PhD is the Eudocia and Sebastiano Papasthatopoulos Endowed Chair in Medicine and an Associate Professor of Medicine and of Epidemiology and Biostatistics. She is the Director of the UCSF Center for Vulnerable Populations at San Francisco General Hospital and the Director of the CTSI K... Show less

Interests
Cardiovascular disease, Health disparities, Health of young adults, Chronic disease prevention

Websites
UCSF Center for Vulnerable Populations at San Francisco General Hospital
UCSF CTSI K Scholars Program

Featured Videos

Publications
1. Coxson PG, Bibbins-Domingo K, Cook NR, Joffres M, Schmidt SM. Response to mortality benefits from U.S. population-wide reduction in sodium consumption: projections from 3 modeling approaches. Hypertension. 2013 Jun; 61(6):e60.
View in: PubMed
2. Chang A, Van Horn L, Jacobs DR, Liu K, Muntner P, Newsome B, Shoham DA, Durazo-Arvizu R, Bibbins-Domingo K, Reis J, Kramer H. Lifestyle-Related Factors, Obesity, and Incident Microalbuminuria: The CARDIA (Coronary Artery Risk Development in Young Adults) Study. Am J Kidney Dis. 2013 Apr 10.
View in: PubMed
3. Moyer VA. Screening for intimate partner violence and abuse of elderly and vulnerable adults: u.s. Preventive services task force recommendation statement. Ann Intern Med. 2013 Mar 19; 158(6):478-86.
View in: PubMed

Kirsten's Networks
Click the "See All" links for more information and interactive visualizations!

Related Concepts
Derived automatically from this person's publications.
Coronary Disease
African Continental Ancestry Group
Heart Failure
Cardiovascular Diseases
European Continental Ancestry Group
» See all (336) concept(s)

Co-Authors
People in Profiles who have published with this person.
Peralta, Carmen
Vittinghoff, Eric
Whooley, Mary
Shlipak, Michael
Pletcher, Mark
» See all (37) people

Related Authors
People who share similar concepts with this person.
Powe, Neil
Massie, Barry
Whooley, Mary
Shlipak, Michael
Dracup, Kathy
» See all (60) people

Same Department
Barnett, Christopher
Moody, Sandra
Nadler, Paul
Tran, Lien
Woeber, Ken
» Search Department

UCSF | chatter

Login to create a UCSF Chatter group right from UCSF Profiles!

Naming Conventions

We abbreviate the RNS “UCSF Profiles” to “Profiles”. We also use the term “profile” without capitalization to denote individual pages describing researchers. Although we describe the users and usage of an RNS, we refer to those who use the site as “visitors” to minimize confusion between our description of those who visit the site versus those who own profiles on the site and can log in to access additional owner-related functions.

Time Period of Analysis

To understand visitors to Profiles and how they evolved over time, we analyzed data from December 2009 (the beta launch of Profiles) through December 2012. We primarily present time course data for the 28 months from the public launch in September 2010 through December 2012 and summary data from the last 6 months analyzed (Months 23-28), from July 2012 through December 2012.

Population Represented in Profiles

Individuals were added mostly en masse and offered the option to opt-out of the system (7 did), or invited to opt-in if their relationship with the university was less formal or if otherwise deemed appropriate (for example, faculty whose primary affiliation is with another institution but have “without-salary” appointments at UCSF). A total of 7297 profiles were created from the launch date to December 31, 2012, at which point there were 5928 active profiles. The difference represents people no longer at UCSF and the 7 who chose to opt out. Individuals added to Profiles by December 2012 included 2691 full time faculty, 356 without-salary faculty, 386 instructors, 2079 trainees (postdoctoral scholars, residents, and fellows), and 1785 other university leaders, researchers, and staff.

Web Analytics

Website usage was measured and analyzed primarily using Google Analytics, which is an industry standard Web analytics measurement framework [23], used by 51% of Fortune 500 companies [24] and 66% of the top 10,000 websites online [25]. We chose Google Analytics for its long-term support, and extensive data filtering and segmentation capabilities—reasons why it has begun to be used for other academic sites as well [26].

We used terminology consistent with that used by Google. We filtered out IP addresses associated with our internal office to remove the effect of internal testing and usage. “Visit” indicated a series of one or more interactions with the website, separated by gaps of no more than 30 minutes between subsequent interactions. “Pages/Visit” represented the number of individual Web pages accessed over the course of a visit. “Time on Site” meant the length of time between the first and last interaction during a visit (time on site is ignored for visits of only one page). “Visitor Location” indicated a visitor’s organizational and/or geographic location as inferred by performing geolocation and network ownership analysis of a user’s IP address. We divided locations into five non-overlapping categories: UCSF, San Francisco, California, US, and World.

“UCSF” included visitors who originated from an IP address associated with the UCSF internal network; 99.73% (114,944/115,258) of these visits originated from facilities in the City of San Francisco. This underestimates UCSF visitors because it excludes some UCSF sites (eg, San Francisco Veterans Administration Hospital, the San Francisco General Hospital, and the Gladstone Institutes) and those using mobile devices or working off-site without a VPN. To help us understand the degree to which we underestimated UCSF visits, we analyzed IP addresses of those who opened emails targeted specifically to UCSF employees (using the tool MailChimp). Of 615 opened emails, 344 (55.9%) were opened from campus IP addresses and 271 (44.1%) from other external IP addresses. Though it may be more common for mobile visitors to read email than access websites, this nevertheless suggests that visits from the UCSF campus network reflect only a portion of visits from all UCSF personnel, and the actual percentages could be up to 79% higher than the numbers we see coming from the campus network. In this paper, however, when we refer to UCSF visitors, we mean only those unambiguously located with the campus network.

“San Francisco” included visitors who originated from the City of San Francisco, other than those from the UCSF internal network. We expect that this included almost all of the remaining UCSF sites that were not otherwise covered by the UCSF network, though their contributions to overall visits are small. San Francisco is also the location most likely to capture traffic from patients of the UCSF Medical Center.

“California” included visitors who originated from California other than those from either San Francisco or the UCSF internal network. This too included some UCSF-related traffic, since many researchers and staff live in the greater Bay Area outside of San Francisco. “US” included visitors who originated from the United States outside of California, and “World” included visitors who originated from sites outside the United States.

Under “Referring Source”, terms included (1) “Search”, visits from search engines (eg, Google, Bing, Yahoo, UCSF.edu search engine), (2) “Website Referrals”, visits from other websites, excluding traffic via search engines, and (3) “Direct or Unknown”, visits from sources that cannot be programmatically identified (eg, user typing in “profiles.ucsf.edu” on their Web browser, clicking a browser bookmark, clicking a link in a desktop email or Twitter client, or following a link from websites that use the secure HTTPS protocol to the non-secure Profiles website). For “New and Returning Visits”, returning visits are those sessions (visits) where the visitor is recognized as having visited the site before via the presence of cookies. Because of the reliance on cookies, this is generally an underestimate of the true value.

Categories of Visitor Actions

We report on “Depth of Visit”, which can be measured as time on the site or as pages per visit (pages/visit). We used pages/visit since that most directly addresses the next steps visitors take after viewing a profile page. A one-page visit might mean that the visitor arrived on a profile page and then stopped using Profiles or clicked a link to some other website (such as PubMed) without returning in the next 30 minutes. Any visit that lasts for 2 or more pages occurs because a visitor clicks on one of three types of links from a profile page. These three categories of links are depicted in Figure 1. The first set of links (A) reveals related researchers or topics either through reinitiating a search (A1) or by clicking on a name, topic, or link that expands to related names or topics (A2). The second link (B) enables editing of one’s own UCSF profile. The third set of links (C) describes Profiles.

Feedback Survey Analysis

Between September 3, 2011 and January 31, 2013, site visitors were asked, “How has Profiles helped you?” The survey appeared on all Profiles pages to all visitors (unless the visitor chose to minimize the survey’s inline pop-up), and 475 comments were submitted. Of these, 5 comments were deleted for having no content or email address, and 1 was deleted as it came from a Profiles developer. The remaining 469 responses were binned into seven categories based on the content and submitter’s email address (if available): (1) Enables Research, (2) Provides Background Information for Clinical Care, (3) Provides Contact Information for Clinical Care, (4) Generally Positive, (5) Negative, (6) Other, and (7) Spam.

Statistics

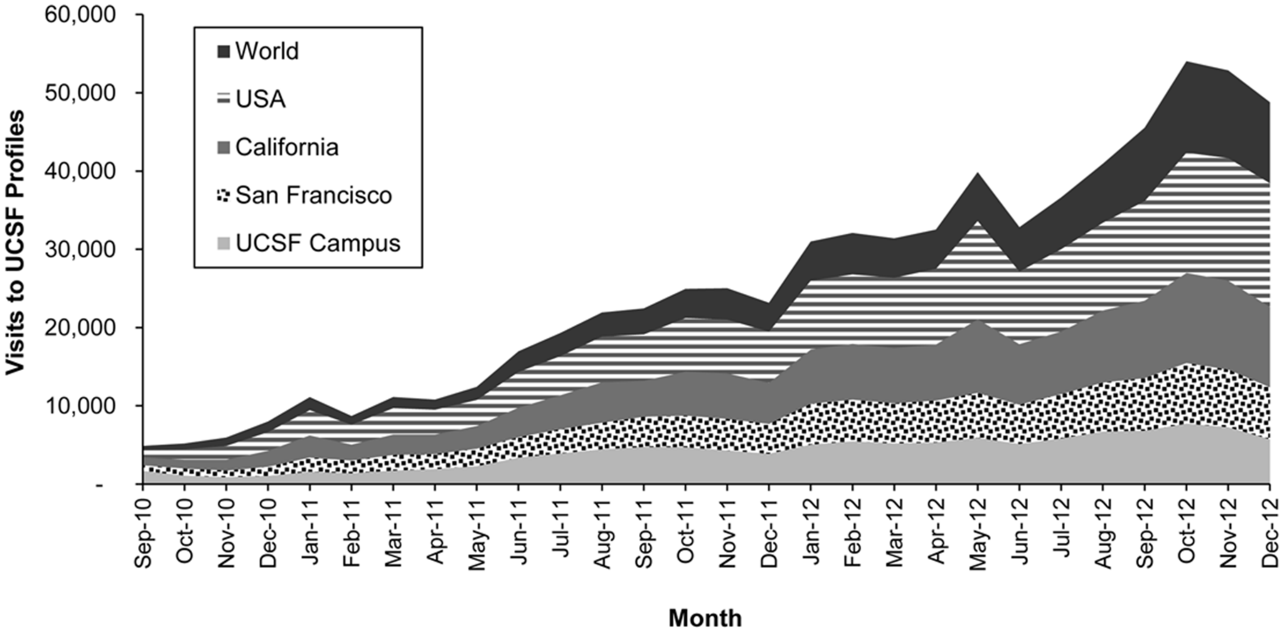
We assessed whether depth of usage of Profiles (measured as pages per visit) differed based on whether a visit was a first-time or returning visit. We also assessed whether UCSF visitors differed in their behavior from other visitors. Unfortunately, Google Analytics does not provide access to raw data describing depth of visit (pages/visit) for each individual visit and as a result statistics were performed on daily averages for each category measured. Daily averages were analyzed for 6 months from July 1 through December 31, 2012, resulting in n=184 days for which average daily depth of visit was calculated for new and returning visitors. A standardization approach was used to control for varying “n” contributing to daily averages. Two-factor analysis of variance (ANOVA) with multiple comparisons was performed on the data. The 2x5 ANOVA assessed whether there was an interaction effect of visitor status (new or returning) with location (UCSF, San Francisco, California, USA, World) in explaining depth of usage of Profiles. All tests were two-sided and a statistical comparison or model was considered significant if $P<.05$. All analyses were performed in Stata v.12.

Results

External Visitors Contribute to Rapid Increases in Visits

Figure 2 describes the pace at which visits to Profiles increased over the 28-month period since launch (September 2010-December 2012). The figure also depicts the institutional and/or geographical location of visitors, ranging from UCSF campus to visitors coming from outside the United States. In the 28 months following launch, traffic to Profiles increased both from on-campus and off-campus sources to an average of 46,000 visits per month as measured for the last 6 months of data analyzed, from July to December 2012. Over this 6-month period, out of a total of 278,570 visits, 40,140 (14.41%) of visits came from UCSF and 238,640 (85.67%) from outside UCSF. As we note in the Methods section, the true percentage of visits from UCSF may be up to 80% higher. Out of the total 278,570 visits, those from outside UCSF can be broken down as 40,438 (14.51%) from San Francisco, 60,256 (21.63%) from California, 81,238 (29.16%) from the United States, and 56,708 (20.36%) from outside the United States. Each category excludes the prior. Visits have continued to increase rapidly; in January 2014, the site received over 85,930 visits, and the last week of January saw more than 3000 visits per weekday.

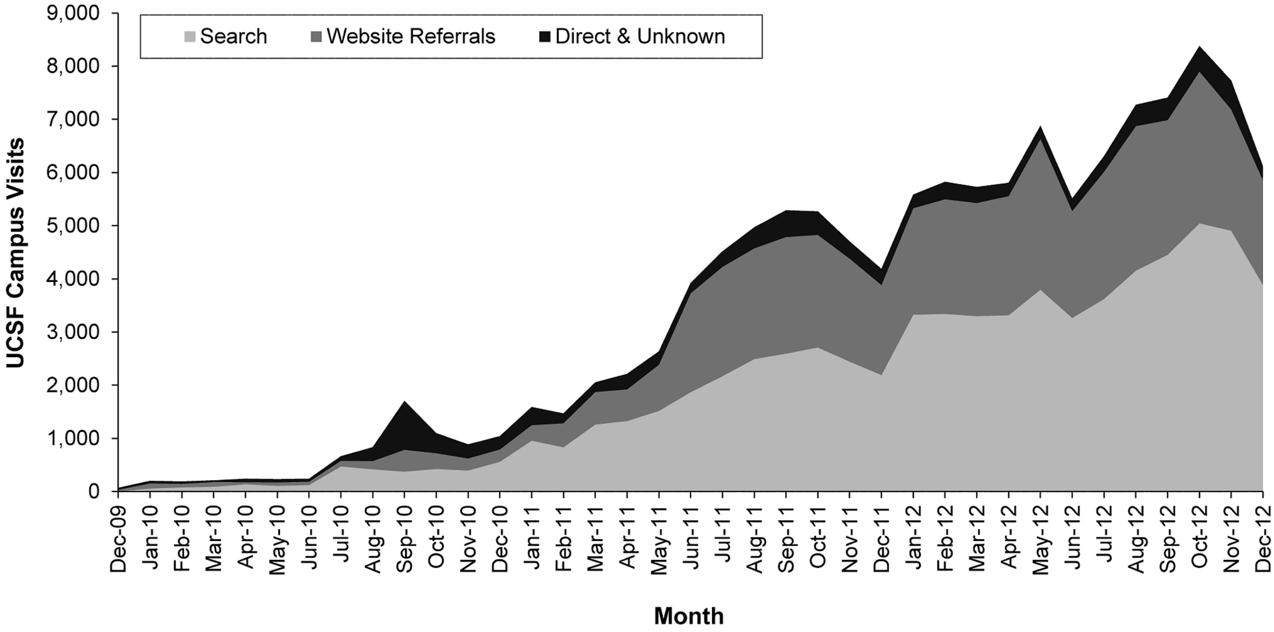
Figure 2. Location of visitors to Profiles for the 28 months since launch.



Search Trumps All Other Sources of Visits

We assessed how visitors arrived at Profiles by analyzing the referring sources of traffic. From July to December 2012 (Months 23-28), search traffic accounted for 208,304 (74.78%) of the 278,570 visits to Profiles, referrals from other websites accounted for 47,861 (17.18%), and direct or unknown sources for 21,926 (7.87%). Google dominated search traffic; 190,170 (91.29%) of the 208,304 visits from a search engine came via Google, while another 7992 (3.84%) came via the search engine on the UCSF website, which is powered by Google and uses the same underlying ranking algorithms. The remaining 10,142 (4.87%) of search-driven traffic (“Other search”) came via sources such as Bing and Yahoo. Visitors from all sources arrived on Profiles via search, but non-UCSF visitors had a greater likelihood of landing on Profiles via a search engine than UCSF visitors (80% for non-UCSF vs 63% for UCSF). The time course of visitor acquisition from search, website referrals, and direct or other sources is shown in Figure 3.

Figure 3. Time course of growth in traffic from various referral sources (peak in direct traffic in September 2010 coincides with launch activities at UCSF).



Visitors Arrive at Profiles Predominantly by Searching for People's Names

We analyzed 250 search queries that landed on Profiles randomly selected from the set of such queries during Months 23-28. In this sample, of all searches that sent visitors to Profiles, 210 (84.0%) included a name, with half of these only including a person's name (42.0%, 105/250). Another large set used the name along with a reference to the university (33.2%, 83/250; eg, <name> ucsf ; <name> university of california san francisco). A smaller set searched for a name along with additional information (8.8%, 22/250; eg, <name> radiation oncology). Finally, 10.0% (25/250) used the name of a paper, and the remaining 6.0% (15/250) included a mix of the name of the university (without other information), URLs, and other data.

Visitors Return in Increasing Numbers

The percentage of visits from returning visitors doubled in the first year (from 1203 of 7039 visits, or 17.09%, in the month of launch, September 2010, to 7235 of 21,924 visits, or 33.0%, in August 2011) but remained constant after that. But as visits continued to increase, this means that the number of visitors who returned to the site continued to increase as well. A higher proportion of UCSF visits included returning visitors (26,758 of 40,140, or 66.66% in Months 23-28) compared to non-UCSF visits (66,105 of 238,430, or 27.73% in Months 23-28). But returning visits continued to increase in absolute numbers for both UCSF and non-UCSF visits. Returning visits are of particular interest not only because they represent a group of visitors who find the tool useful, but, as we show below, these visitors also tend to use the tool in more depth.

Subset of Visitors Continue Beyond Profile Page to Explore Related People and Topics

We analyzed the distribution of visits by depth of visit for the period from Months 23-28. As Figure 4 shows, the majority of visits during Months 23-28 resulted in the viewing of only 1 page (81.15%, 225,750/278,184) with the remaining 18.85% (52,434/278,184) extending for 2 or more pages. Visits that went 2 or more pages utilized links shown in Figure 1 in categories A, B, and C. Even though most visits were only 1 page deep, a significant number of visits included multi-page browsing. For example, in January 2013, the site received 2641 visits (or >85 visits a day) where visitors browsed the site for 5 or more pages. To provide a flavor of the average time spent on site during this period, an average 2-page visit lasted for 4 minutes and 8 seconds, a 10-page visit lasted for 12 minutes and 8 seconds, and a 20-page visit lasted for 17 minutes and 34 seconds.

To further understand the distribution of actions taken by visitors immediately after they landed on a profile page, we evaluated actions taken on the first page of 1024 visits to profile pages randomly selected from all visits during Months 23-28. The results are described in Table 1. In this sample, 76.27% (781/1024) either left the site after viewing 1 page or took a path that Google was unable to identify. Visits that left the site included those that clicked on links for individual publications. The remaining 23.73% (243/1024) would be counted as visits with 2 or more pages viewed. These can be broken down as 10.84% (111/1024) clicked to see other related people or concepts (eg, as presented by the novel networking elements of the RNS, Category 1A and B in Figure 1), 2.34% (24/1024) clicked back to the list of search results and clicked on another person, and 7.71% (79/1024) initiated a new search. A small number went on to edit their own profile (0.98%, 10/1024, Category B) or clicked on a site-wide informational link (1.86%, 19/1024, Category C).

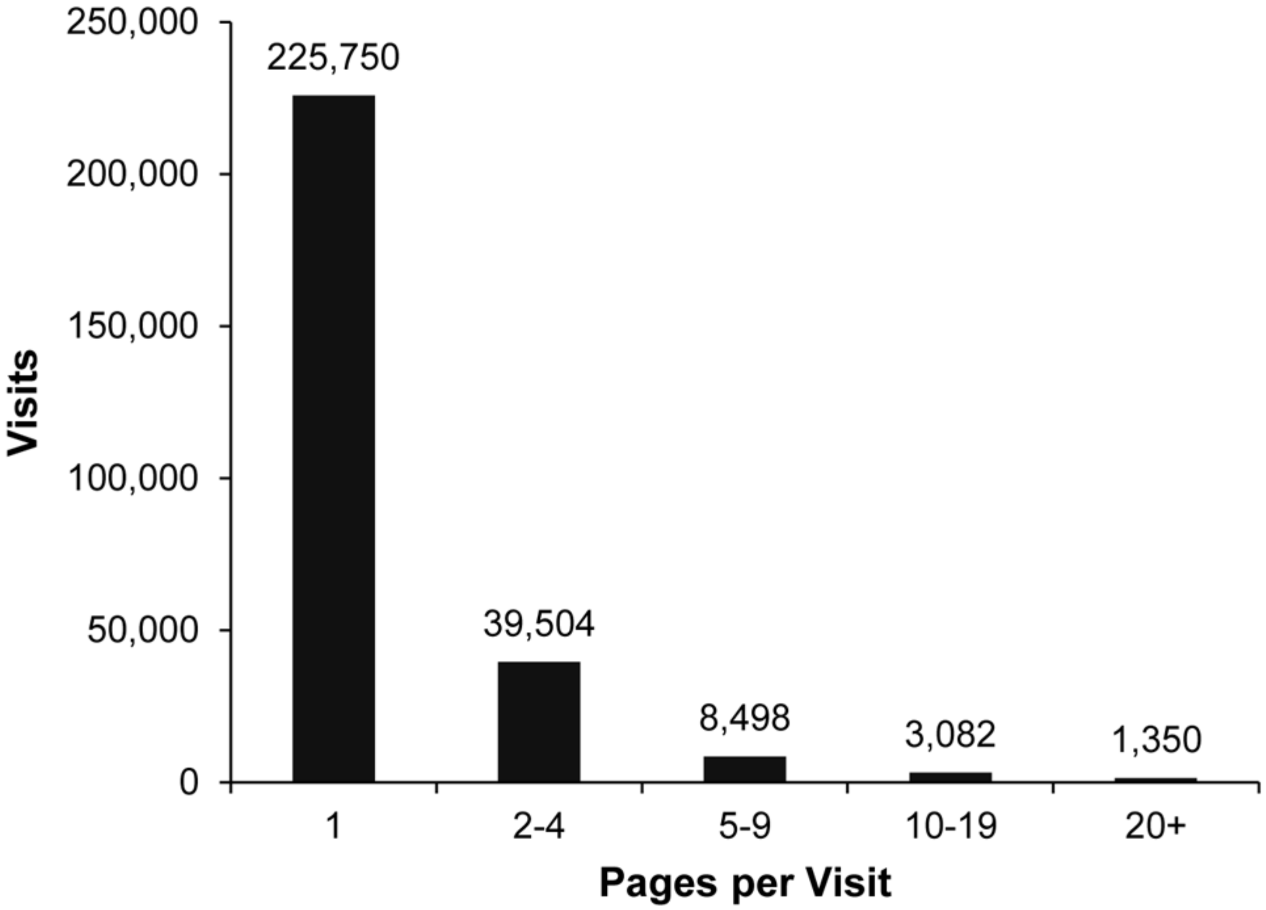
Table 1. Actions taken from a profile page (n=1024).[a]

Action category	n (%)	Descriptive data
Leave (Exit)	706 (68.95)	Profile page is last page user looked at before leaving the website
More about that person or topic's relationships [b]		
	20 (1.95)	Clicks on list of all Co-Authors
	12 (1.17)	Clicks on list of all Similar People
	5 (0.49)	Clicks on list of all Keywords for that person
Another person [b]		
	74 (7.23)	Clicks to another user's profile (eg, by clicking on a Co-Author, Similar Person, etc, link)
	24 (2.34)	Clicks back to list of search results, then clicks on another person
New Search[b]	79 (7.71)	Runs a new search (eg, by entering text in mini search box)
Site-wide navigation		
	13 (1.27)	Clicks to homepage
	6 (0.59)	Clicks on "How Profiles Works" page
Edit Profile	10 (0.98)	Clicks on edit profile link
Unknown	75 (7.32)	Unknown or unrecorded

[a] Analysis of user behavior on the initial landing page for 1024 visits to profile pages randomly selected from all visits during Months 23-28.

[b] Paths to "networked links" that connect the initial profile page to related people or topics.

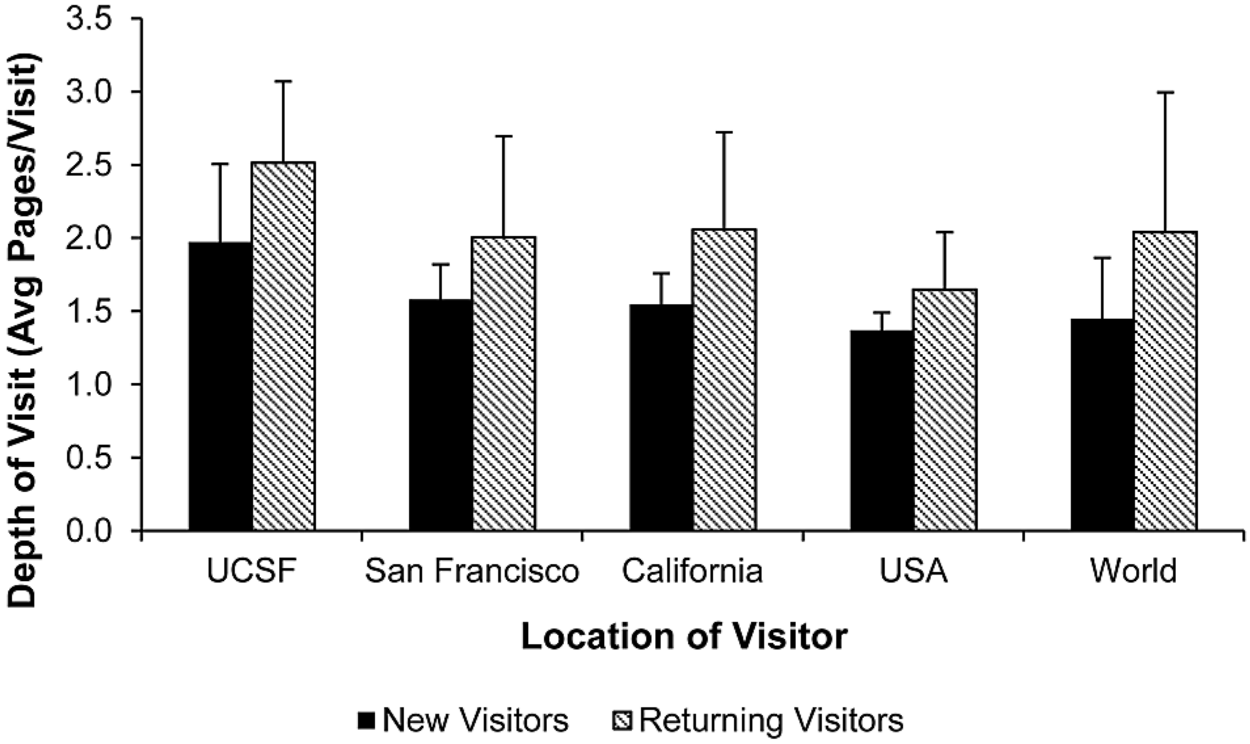
Figure 4. Distribution of visitors by depth of visit (pages/visit) during Months 23-28 from launch.



Returning Visitors Use the Site More Deeply Than First-Time Visitors

Figure 5 shows the depth of visit as average pages/visit for new and returning visitors coming from different locations. Returning visitors with an average 2.103 pages per visit went deeper than new visitors with an average 1.496 pages per visit, regardless of the location they came from ($P<.001$, standardized mean difference of 1.099). Depth of visit also depended on the location that visitors came from with UCSF visits resulting in the deepest visits ($P<.001$). Although visit numbers were lower on weekends, the depth of visit did not differ for any day of the week.

Figure 5. Depth of visit (pages/visit) for new and returning visitors. The figure describes visitors originating from different locations during Months 23-28 from launch. Error bars show 1 Standard Deviation.



Users Provide Direct Feedback Revealing Broad Utility of RNS

In September 2011, we implemented an online user feedback form that asked visitors, “How has Profiles helped you?” During the 16 months after implementation (through Jan 2013), 469 submissions were received from a total of 670,000 visits, with 0.07% of all visits converted to a submission. Overall, 284 (60.5%, 284/469) comments describing positive outcomes were submitted. We found 58 (12.4%, 58/469) described how Profiles enabled research, 91 (19.4%, 91/469) described how Profiles enabled clinical care by providing contact information, 18 (3.8%, 18/469) described how Profiles enabled clinical care by providing research-related information on clinicians, and 117 (24.9%, 117/469) were positive without specific details. In addition, 118 (25.2%, 118/469) categorized as “Other” primarily requested help with the tool, 3 (0.6%, 3/469) were spam, and 64 (13.6%, 64/469) were “Negative”. Of the “Negative” comments, 29 (45%) out of 64 left negative comments reflecting their unmet expectations of a traditional directory listing (such as organization by department or sub-specialty), while 21 (33%) simply said “no”, and the remainder noted a negative comment related to technical difficulties. We show a sample of the comments, categorized as “Enabling Research”, in Textbox 1. Visitors self-reported that they found collaborators, identified new research problems, and improved administrative processes.

Textbox 1. Examples of user feedback that described Profile’s positive impact on research. The question, “How has Profiles helped you?” was asked on a pop-up survey automatically presented to all visitors. The survey was deployed during Months 13-29, from September 2011 through January 2013. These are 13 illustrative quotes of user feedback from the 58 user comments that were categorized as “enabling research”.

I am hoping that it just helped me find a mentor...

Found speakers for grand rounds at SFGH pediatrics.

Great resource for finding potential research collaborators and for PhD dissertation committees.

It is wonderful to find the current publications. My faculty rarely tell me. I need it for our annual report and for some grant applications.

This profile allowed me to review the background and experience of an author from a journal article I was reading.

Research information about scientists for my biomedical foundation.

> *Highlights/generates awareness about science accomplishments beyond my immediate sphere of research but very stimulating and interesting.*
>
> *Gives me a profile of a potential contributor to a book that is under consideration, helps to locate people in a specific field. Important service!*
>
> *Making your remarkable research available to me and my translational research projects is invaluable. Thank you so very much.*
>
> *Trying to find interesting colleagues to meet with during my stay at UCSF.*
>
> *Quite useful in introducing UCSF colleagues to others outside the university. Also a quick way to get a feel for the interests of people I don't know.*
>
> *To find researchers with common interests.*
>
> *Contact a faculty member recommended by an associate. Initiate a sponsored project.*

Discussion

Principal Findings

Research networking systems automatically generate and publically display profiles and networks of researchers. Within 2.5 years, our system, UCSF Profiles, achieved one-seventh of the monthly visit rate of the longstanding main university website, with a continually increasing trend. Visitors came from diverse locations beyond the institution. Responses to an online survey suggest a broad range of benefits supporting the research and clinical mission. Returning visitors are among those that appear to use the system to identify new collaborators or research topics. And, with the continually increasing visits arising from public searches for individual names, UCSF Profiles is evolving into a new kind of online gateway for the university.

Assessing and Improving the Usage of Research Networking Systems

Schleyer et al propose a comprehensive framework to advance the study of research networking systems [26] within which this work would fall under their "evaluation" axis. They define RNSs as systems that enable collaboration and develop a framework to test that specific hypothesis. In contrast, we take an empirical approach to understanding the role of an RNS at an academic institution and assess our customers and their online behavior just as a commercial vendor would. The profile page of a person of interest was the main entry point for visitors. For those that left the page (nearly 80%), we cannot distinguish between those who began a new task from those who continued to explore the original person of interest but had to leave the site (eg, by going to a publication in PubMed or clicking on a YouTube video of a talk). For those who continued on the site, almost all of the visitors (roughly 20%) clicked on links of people who were not the original person of interest but were displayed as related people (coauthors or similar people), or clicked on topics exposed by the tool, which in turn led to a new listing of people. We interpret our results as showing that at least one-fifth of visits initiated the process of exploring a related researcher or topic, someone or something the visitor had not considered in their initial search. But we cannot conclude how frequently they achieved their goal. On the other hand, visits continue to increase at a rapid pace and visitors return and stay longer in increasing numbers, suggesting a positive outcome and perceived utility for a growing segment. Specific responses to our online survey provide anecdotal evidence of impact and define a framework for systematic assessment. In addition to identifying collaborators and building teams, based on the responses, the framework should assess impact on creating efficiencies in research administration, enabling broader research functions such as mentoring, and advancing the clinical mission.

Although the initial goal of Profiles was to enable collaboration by making it easier for researchers to find partners and build teams, the most striking aspect of its deployment was how rapidly visitors were drawn from both within and outside the university. In January 2013, visits to UCSF Profiles were one-seventh of the overall visits to the established UCSF campus website and, at the time of publication, visits had caught up to almost one-fifth (18%) of that traffic. We attribute several factors to this increase in visits. As Figure 3 describes, increases in visits came both from increased website referrals and from significant increases in visitors coming from search engines. Most of the increases in visits from website referrals reflect increased numbers of UCSF visitors finding Profiles through campus websites. We initiated a high-level partnership with the University Relations office and developed a coordinated strategy to promote university faculty and research. A key element was standardizing links for faculty who were being publicized so they always pointed to their UCSF Profiles page. We also realized that the directory was a significant point of entry for the university. A partnership again allowed us to embed links to profiles in the university directory. Many other interventions, including a data federation strategy where departments could use publication and other feeds from Profiles to enhance their own websites, brought goodwill and adoption across campus, but contributed only a small proportion of visits overall. This was because, as we show, visitors from search engines dominated all sources, which in turn was enabled through a comprehensive search engine optimization strategy implemented early in the deployment of Profiles. While details of search engine optimization techniques are out of scope for this paper, broadly, we encouraged links to UCSF Profiles for a variety of on- and off-campus websites, copyedited HTML page titles and page descriptions, implemented HTML people data microformats, simplified URLs to "profiles.ucsf.edu / firstname.lastname", implemented a sitemap, and cleaned up redirects and error page HTTP codes.

In efforts to understand the rapid pace of acquiring visitors, we also found that most visits to Profiles came from visitors searching for a name. As visits continue to increase, Profiles continues to capture more of this type of visitor. Where might these Internet users have landed had they not found Profiles? Perhaps not the primary university website—compared to more than 70% of visits arriving via search to Profiles, only 25% of visits come via search to the university website. But, most researchers have multiple affiliated websites, some official such as departmental, laboratory, or clinical, and some not, such as news articles, and some related to other organizations such as a journal or corporation. The visits being captured by Profiles were otherwise likely to be dispersed among these various possibilities. Instead, they are now being aggregated and captured by the university RNS. And, with its search-optimized interface and increasing number of personally-curated elements, the site is also becoming the link of choice for external media. When Gurpreet Dhaliwal was profiled in the *New York Times* as a masterful clinical diagnostician, the newspaper linked to his profile in UCSF Profiles [27]; a variety of other media outlets have also linked to individual profiles.

People First: A New Gateway Exposes the Strengths of the University

An assessment of the searches that bring visitors to Profiles does not describe the universe of searches. Still, of those brought to Profiles via search engines, we found that 84% searched for individuals' names with or without accompanying search terms including the institution's name. But regardless of whether a visitor to Profiles is aware of the institutional affiliation of the person for whom they search, by the time they land on a profile page they are in effect being exposed to the institution. And when they continue on the site, they are introduced to the university through the unique lens of expertise and intellectual networks, not schools, departments, or administrative organization. That Internet users search for people at a university independent of their interest in the university itself makes sense even if we may not have predicted the volume of this trend. Public affairs offices have long understood this—promoting the university through news stories about individual faculty. But an RNS such as Profiles (among others [4]) is explicitly designed to promote people, their intellectual outputs, and networks at scale. Thus, an RNS presents data in a format optimized for the most ubiquitous consumer tool (search engines) delivering content optimized around consumer interests (names of individuals). As a result, the institutional RNS is evolving into a new online gateway for the university, providing a discoverable interface to the intellectual capital of the institution: its people, knowledge, and networks. With this new frame, we are reassessing the positioning of this publicly accessible enterprise system within the fabric of the university. Further integration of the RNS into the university public relations office, into the Executive Vice Chancellor and Provost's office, and indeed into the medical center's marketing and communications arm are relevant next steps.

Acknowledgments

We thank Clay Johnston and Dan Lowenstein for providing inspiration and indispensable advice, David Arrington for facilitating the assessment of Profiles in the context of university websites, Isabel Allen for help with statistics, and Sarah Nelson for superb image production contributions. This publication was supported by NSF Award 1160906 and the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), through UCSF-CTSI Grant Number UL1 TR000004. The Profiles Research Networking Software is supported by NCATS NIH Award 1UL1 TR001102-01. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

References

1. Levy A. Bill Gates joins $35 million funding in startup ResearchGate. URL: http://www.bloomberg.com/news/2013-06-04/bill-gates-joins-35-million-investment-in-startup-researchgate.html [accessed 2013-11-13] [WebCite Cache ID 6L6Qt2d9g]
2. Gewin V. Collaboration: Social networking seeks critical mass. Nature 2010 Dec 16;468(7326):993-994. [doi: 10.1038/nj7326-993a]
3. Zerhouni EA. Translational research: moving discovery to practice. Clin Pharmacol Ther 2007 Jan;81(1):126-128. [doi: 10.1038/sj.clpt.6100029] [Medline: 17186011]
4. DIRECT2Experts Network. 2013. URL: http://direct2experts.org/ [accessed 2013-11-13] [WebCite Cache ID 6L6R0oWfN]
5. Weber GM, Barnett W, Conlon M, Eichmann D, Kibbe W, Falk-Krzesinski H, Direct2Experts Collaboration. Direct2Experts: a pilot national network to demonstrate interoperability among research-networking platforms. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i157-i160 [FREE Full text] [doi: 10.1136/amiajnl-2011-000200] [Medline: 22037890]
6. Northwestern Scholars. URL: http://www.scholars.northwestern.edu/ [accessed 2013-11-13] [WebCite Cache ID 6L6RS2zu8]
7. Vivo at University of Florida. URL: http://vivo.ufl.edu/ [accessed 2013-11-13] [WebCite Cache ID 6L6RVObfL]
8. Schleyer T, Butler BS, Song M, Spallek H. Conceptualizing and Advancing Research Networking Systems. ACM Trans Comput Hum Interact 2012 Mar 1;19(1):1 [FREE Full text] [doi: 10.1145/2147783.2147785] [Medline: 24376309]

9. DiMicco J, Millen DR, Geyer W, Dugan C, Brownholtz B, Muller M. Motivations for social networking at work. 2008 Presented at: Proceedings of the ACM Conference on Computer Supported Cooperative Work; 2008; San Diego, CA, USA p. 711-720.
10. Riemer K, Scifleet P. Enterprise social networking in knowledge-intensive work practices: A case study in a professional service firm. In: ACIS. 2012 Dec Presented at: Australasian Conference on Information Systems; December 3-5, 2012; Geelong, Australia p. 1-12.
11. Dorit N, Benbasat I, Wand Y. Who knows what?.: Wall Street Journal / Sloan Management Review; 2009 Oct 26. URL: http://online.wsj.com/news/articles/SB10001424052970203946904574302032097910314 [accessed 2013-11-13] [WebCite Cache ID 6L6RZ4sOc]
12. Jones BF, Wuchty S, Uzzi B. Multi-university research teams: shifting impact, geography, and stratification in science. Science 2008 Nov 21;322(5905):1259-1262 [FREE Full text] [doi: 10.1126/science.1158357] [Medline: 18845711]
13. Richter A, Riemer K. Corporate social networking sites: modes of use and appropriation through co-evolution. 2009 Presented at: 20th Australasian Conference on Information Systems; 2009; Melbourne, Australia.
14. Perer A, Guy I, Uziel E, Ronen I, Jacovi M. The longitudinal use of SaNDVis: visual social network analytics in the enterprise. IEEE Trans Vis Comput Graph 2013 Jul;19(7):1095-1108. [doi: 10.1109/TVCG.2012.322] [Medline: 23661007]
15. Lampe C, Ellison N, Steinfield C. A face(book) in the crowd: Social searching vs social browsing. 2006 Nov Presented at: 20th Anniversary Conference on Computer Supported Cooperative Work; 2006; Banff, Alberta, Canada p. 167-170.
16. Ellison NB, Steinfield C, Lampe C. The benefits of Facebook friends: social capital and college students use of online social network sites. Journal of Computer-Mediated Communication 2007;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
17. Boland MR, Trembowelski S, Bakken S, Weng C. An initial log analysis of usage patterns on a research networking system. Clin Transl Sci 2012 Aug;5(4):340-347 [FREE Full text] [doi: 10.1111/j.1752-8062.2012.00418.x] [Medline: 22883612]
18. Bhavnani SK, Warden M, Zheng K, Hill M, Athey BD. Researchers' needs for resource discovery and collaboration tools: a qualitative investigation of translational scientists. J Med Internet Res 2012;14(3):e75 [FREE Full text] [doi: 10.2196/jmir.1905] [Medline: 22668750]
19. Schleyer T, Spallek H, Butler BS, Subramanian S, Weiss D, Poythress ML, et al. Facebook for scientists: requirements and services for optimizing how scientific collaborations are established. J Med Internet Res 2008;10(3):e24 [FREE Full text] [doi: 10.2196/jmir.1047] [Medline: 18701421]
20. 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]
21. Deep N. 200 million members!. URL: http://blog.linkedin.com/2013/01/09/linkedin-200-million/ [accessed 2013-11-13] [WebCite Cache ID 6L6WXAPsR]
22. Profiles research networking software. URL: http://profiles.catalyst.harvard.edu/ [accessed 2013-11-13] [WebCite Cache ID 6L6WlSyrM]
23. Clifton B. Advanced web metrics with Google Analytics.: Wiley.com; 2012. URL: http://www.wiley.com/WileyCDA/ [accessed 2013-11-13] [WebCite Cache ID 6L6hqP75b]
24. Farina C. e-nor Blog. 2012 Jul 09. 51% of Fortune 500 companies use Google Analytics URL: http://www.e-nor.com/blog/web-analytics/51-of-fortune-500-companies-use-google-analytics [accessed 2013-11-13] [WebCite Cache ID 6L6Wsu0Kw]
25. BuiltWith. BuiltWith Web Technology Usage Trends. Google Analytics Usage Statistics URL: http://trends.builtwith.com/analytics/Google-Analytics [accessed 2013-11-13] [WebCite Cache ID 6L6bdqnSj]
26. Fagan JC. The Suitability of Web Analytics Key Performance Indicators in the Academic Library Environment. The Journal of Academic Librarianship 2013 Aug:1. [doi: 10.1016/j.acalib.2013.06.005]
27. Hafner K. For Second Opinion, Consult a Computer?.: New York Times; 2012 Dec 3. URL: http://www.nytimes.com/2012/12/04/health/quest-to-eliminate-diagnostic-lapses.html?pagewanted=all [WebCite Cache ID 6L6WQNnCo]

Abbreviations

ANOVA: analysis of variance
RNS: research networking system
UCSF: University of California, San Francisco

Edited by G Eysenbach; submitted 26.11.13; peer-reviewed by M Rethlefsen, P Garaizar; comments to author 16.12.13; revised version received 10.01.14; accepted 19.01.14; published 07.02.14.

Please cite as:
Kahlon M, Yuan L, Daigre J, Meeks E, Nelson K, Piontkowski C, Reuter K, Sak R, Turner B, Weber GM, Chatterjee A
The Use and Significance of a Research Networking System
J Med Internet Res 2014;16(2):e46
URL: http://www.jmir.org/2014/2/e46/
doi:10.2196/jmir.3137
PMID:24509520

Original Paper

Did You Hear the One About the Doctor? An Examination of Doctor Jokes Posted on Facebook

Matthew A Davis[1,2], MPH; Carol Sue Haney[3]; William B Weeks[1], MBA, MD; Brenda E Sirovich[1,4,5], MS, MD; Denise L Anthony[6], PhD

[1]The Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH, United States
[2]Institute for Quantitative Biomedical Sciences Graduate Program, Dartmouth College, Hanover, NH, United States
[3]Toluna, Wilton, CT, United States
[4]Outcomes Group, VA Medical Center, White River Junction, VT, United States
[5]Geisel School of Medicine at Dartmouth, Hanover, NH, United States
[6]Department of Sociology, Dartmouth College, Hanover, NH, United States

Corresponding Author:
Matthew A Davis, MPH
The Dartmouth Institute for Health Policy and Clinical Practice
35 Centerra Parkway
Lebanon, NH, 03766
United States
Phone: 1 (603) 653 0842
Fax: 1 (603) 653 0896
Email: matthew.a.davis@dartmouth.edu

Abstract

Background: Social networking sites such as Facebook have become immensely popular in recent years and present a unique opportunity for researchers to eavesdrop on the collective conversation of current societal issues.

Objective: We sought to explore doctor-related humor by examining doctor jokes posted on Facebook.

Methods: We performed a cross-sectional study of 33,326 monitored Facebook users, 263 (0.79%) of whom posted a joke that referenced doctors on their Facebook wall during a 6-month observation period (December 15, 2010 to June 16, 2011). We compared characteristics of so-called jokers to nonjokers and identified the characteristics of jokes that predicted joke success measured by having elicited at least one electronic laugh (eg, an LOL or "laughing out loud") as well as the total number of Facebook "likes" the joke received.

Results: Jokers told 156 unique doctor jokes and were the same age as nonjokers but had larger social networks (median Facebook friends 227 vs 132, *P*<.001) and were more likely to be divorced, separated, or widowed (*P*<.01). In 39.7% (62/156) of unique jokes, the joke was at the expense of doctors. Jokes at the expense of doctors compared to jokes not at the expense of doctors tended to be more successful in eliciting an electronic laugh (46.5% vs 37.3%), although the association was statistically insignificant. In our adjusted models, jokes that were based on current events received considerably more Facebook likes (rate ratio [RR] 2.36, 95% CI 0.97-5.74).

Conclusions: This study provides insight into the use of social networking sites for research pertaining to health and medicine, including the world of doctor-related humor.

(J Med Internet Res 2014;16(2):e41) doi:10.2196/jmir.2797

KEYWORDS

physicians; physician-patient relations; social networking; humor

Introduction

Laughter is the best medicine, as the saying goes. Sociologists identify humor as a social phenomenon embedded in interaction [1] that can affirm conventional views about the world [2], highlight status differences within or outside of a group [3,4], enhance social bonds [5,6], or relieve stress [7,8]. It may be the latter effect that underlies the saying about laughter as medicine, but joking about medical care or practitioners may have other implications. Scholars have worried about the declining status and authority of physicians for the past 3 decades or more [9-11]. Although doctor jokes have been around since ancient times (eg, McDonald [12]), joking about doctors may contribute to this decline. Other evidence suggests declining career satisfaction among physicians [13,14] as a possible result of this declining status.

Although it may be difficult to know if doctors think doctor jokes are funny, we do know that generally people tell jokes with the intention of amusing others [1] and this may be accompanied by a complex range of motivations, from the expression of disagreement or dissatisfaction to an indication of endearment and friendship [15]. Freud argued that humor was a socially acceptable form of aggression in modern life, particularly when directed toward high status or powerful others [16]. Research suggests that people use and appreciate humor when the target of the joke is from a group different than their own [3,4], which may both enhance cohesion within a group [5] and relieve tension or stress [17] producing the stated “medicinal” effect of laughter. Given the relatively high status of physicians in society, it makes sense they would be the target of jokes. In addition, because Americans are generally dissatisfied with their health care system [18], joking about doctors and medicine may provide a socially acceptable way to express that dissatisfaction. Joking about medicine is also likely to be an important coping mechanism for patients facing serious illness [19].

Other evidence suggests that higher status group members or those who aspire to higher status are more likely to engage in joking behavior, particularly in status-differentiated groups [20]. In addition, there is a relationship between whether the joker himself laughs first, the number of “audience” members, and how much others laugh at the joke [21]. According to Glenn [21], in groups of 2 or larger it is not typical for the joker to laugh first (unlike in dyads).

Although we seem to know more than one might think about joking and jokers, there is relatively little research on humor in spontaneous conversation [20-22]. This is likely because of the difficulty in capturing spontaneous conversations for analysis. Today, however, a great deal of social interaction occurs online in social networking sites, such as Facebook. Currently Facebook has 874 million active users worldwide, communicating in 70 different languages [23]. Although written Facebook conversations are not the same as in-person interactions, we sought to eavesdrop on casual interactions occurring on Facebook to examine jokes about doctors. Therefore, we performed the first study of social networking site conversations pertaining to health and medicine to examine the prevalence, characteristics, and success of doctor jokes posted on Facebook.

Methods

Ethics Approval

The Committee for the Protection of Human Subjects at Dartmouth College reviewed the study protocol and granted this study an expedited Institutional Board Review.

Sample

To obtain information on doctor jokes posted on Facebook, we used data from the Harris Interactive Research Lifestreaming Panel. Upon agreeing to become a Lifestreaming Panel participant, these individuals gave Harris permission to record their private conversations on social networking sites such as Facebook and Twitter. Although both Facebook and Twitter could potentially be used to study jokes about doctors, we were granted access specifically to the Lifestreaming Panel’s Facebook data. Lifestreaming Panel participants are paid the sum of US $1.00 initially, and they are then eligible to participate in future surveys and activities in which they can receive other payments. Harris collects data from Lifestreaming Panel participants’ social networking sites as well as information on the participants’ characteristics. Harris can search Lifestreaming Panel participants’ Facebook walls using algorithms based on keywords. As of June 2011 (the time of our data collection), there were 33,326 adult Lifestreaming Panel participants.

On Facebook, each user creates a profile and has a personal “wall,” a place for conversing with others. For the purposes of this study, we defined a *conversation thread* as starting with a root post on a user’s wall by either the user or by the user’s “friend,” and including others’ follow-up comments. A fictional example of a conversation thread in which a doctor joke is mentioned can be found in Figure 1 (note: the authors are used to represent Facebook users to protect the identify of study participants).

Figure 1. Fictional example of conversation thread on a Facebook wall.

Matthew Davis
Here's one for you guys. . .
I asked my doctor, "How do you determine whether or not an older person should be put in an old age home?" "Well," he said, "we fill up a bathtub, then we offer a teaspoon, a teacup and a bucket to the person to empty the bath tub." "Oh, I understand," I said. "A normal person would use the bucket because it is biggest." "No" he said. "A normal person would pull the plug. Do you want a bed near the window?"
Like · Comment · Share · 3 hours ago

Carol Sue Haney, Bill Weeks, Brenda Sirovich and Denise Anthony like this.

Carol Sue Haney
Good one!
2 hours ago · Like

Bill Weeks
I'll have to remember that one.
44 minutes ago · Like

Brenda Sirovich
2 much. ROTFL!!
30 minutes ago · Like

Denise Anthony
Goofball.
a few seconds ago · Like

Identification of Doctor Jokes

Using Harris Interactive Research Lifestreaming Panel Facebook data from December 15, 2010 to June 16, 2011, we identified 30,075 individual posts that included the term "doctor(s)" (Figure 2). Because posts were identified first by the keyword doctor(s) and then attached to full Facebook wall threads, 60 doctor jokes were made by Facebook users not part of the Lifestreaming Panel (these individuals posted on a Lifestreaming Panel participant's Facebook wall) and were therefore excluded from our study.

To identify jokes, 2 independent coders read each post that included the term doctor(s) in search of jokes. For practical reasons, coders were asked to only identify what they considered to be canned jokes. A typical canned joke contains an introduction followed by a punch line, and can stand alone as a humorous statement, free from context [24]. Coders identified 442 posts that contained potential canned jokes. Differences in coding were reconciled by a third reviewer. Because Cohen's kappa statistic can underestimate the interreviewer agreement when analyzing rare events, we also calculated the positive and negative agreement between reviewers [25,26]. The kappa statistic, positive, and negative agreement for initial agreement between reviewers was 0.62, 0.65, and 0.99, respectively (Figure 2).

Three study investigators then read all 442 potential jokes and excluded those that were not canned jokes or that contained multiple embedded jokes. Differences were discussed and resolved by consensus. The process resulted in the identification of 321 posted doctor jokes that represented 156 unique canned jokes (some jokes were repeated). Selected examples of the jokes identified can be found in Table 1. We included doctor jokes in our sample whether or not the joke received follow-up posts from the social network.

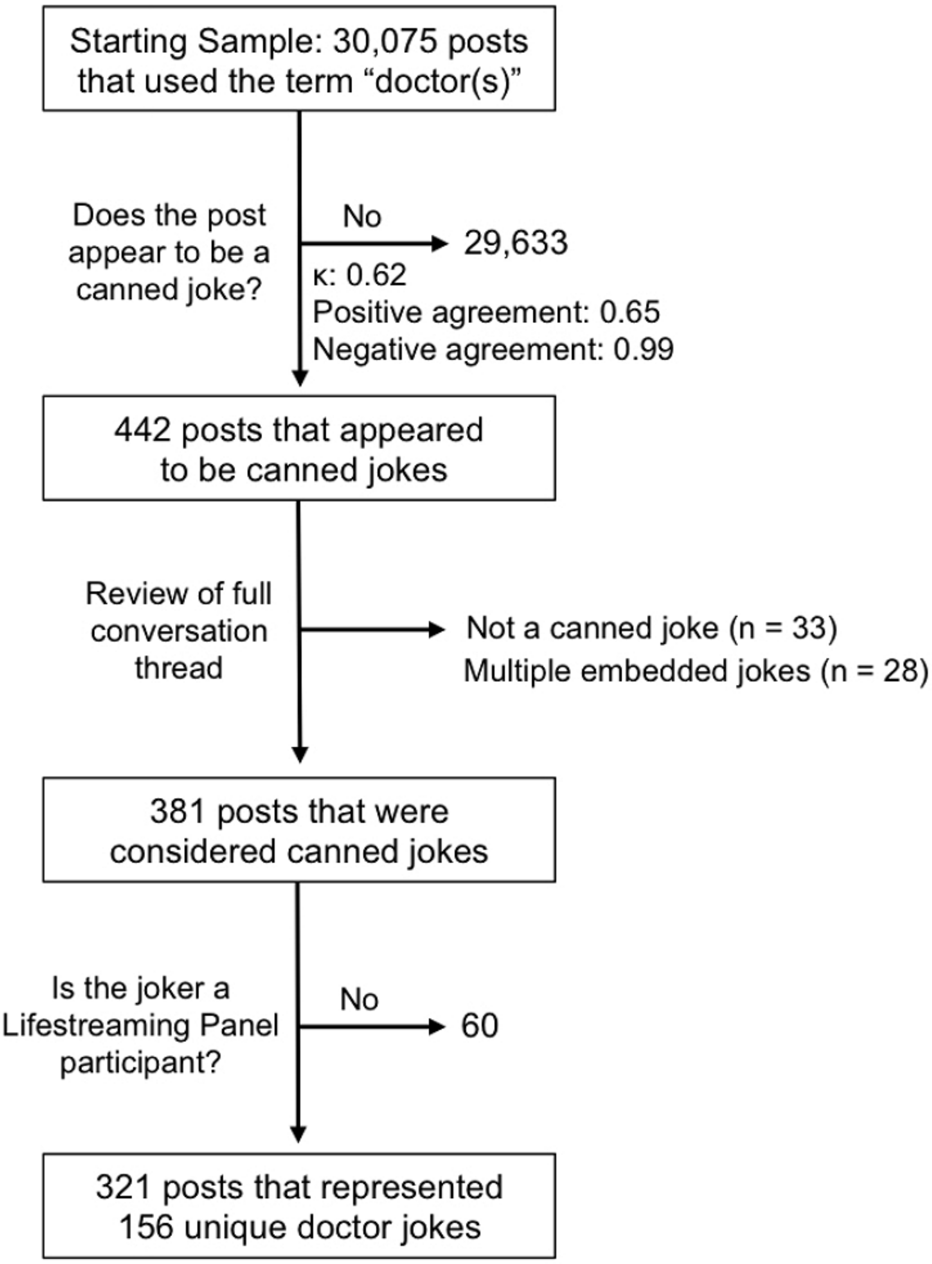
Table 1. Selected examples of Facebook jokes made at the expense of doctors.[a]

Example number	Joke[b]
Medical doctors or the medical profession	
(1)	Doctor, you told me I have a month to live and then you sent me a bill for $1000! I can't pay that before the end of the month! Okay, says the doctor, you have 6 months to live ;)
(2)	Here's a question for you: What do you call a doctor who finishes last in his medical school class? Answer: Doctor.
(3)	An old preacher was dying. He sent a message for his doctor and his lawyer to come. When they arrived, the preacher held out his hands and motioned for them to sit, one on each side of his bed. The preacher grasped their hands, sighed contentedly, smiled, and stared at the ceiling. For a time, no one said anything. Both the doctor and lawyer were touched and flattered that the preacher would ask them to be with him during his final moments. They were also puzzled; the preacher had never given them any indication that he particularly liked either of them. They both remembered his many long, uncomfortable sermons about greed, covetousness, and avaricious behavior that made them squirm in their seats. Finally, the doctor said, "Preacher, why did you ask us to come?" The old preacher mustered up his strength, then said weakly, "Jesus died between 2 thieves and that's how I want to go."
(4)	Frank is recovering from day surgery when a nurse asks him how he is feeling. "I'm fine but I didn't like the 4-letter-word the doctor used in surgery," he answered. "What did he say?" asked the nurse. "Oops!"
(5)	Three out of 4 doctors recommend another doctor.
Health care system	
(1)	Two patients limp into 2 different medical clinics with the same complaint. Both have trouble walking and appear to require a hip replacement. The first patient is examined within the hour, is x-rayed the same day, and has a time booked for surgery the following week. The second calls his family doctor after waiting 3 weeks for an appointment, then waits 8 weeks to see a specialist, then gets an x-ray, which isn't reviewed for another week, and finally has his surgery scheduled for a month from then. Why the different treatment for the 2 patients? The first is a Golden Retriever. The second is a senior citizen :)
(2)	A doctor says to a patient, "We've run every test we can think of and the results show you're out of money."
Poking fun at doctor's advice or the doctor-patient relationship	
(1)	My doctor told me to walk 5 kilometers a day. It's been 5 days and I am 25 kilometers from home and don't know how to get back! Haha.
(2)	So the doctor says, "Take your clothes off and stick your tongue out the window." I asked him, "What will that do?" The doctor says, "I'm mad at my neighbor."
(3)	A woman told her doctor, "I've got a bad back." The doctor said, "It's just old age." The woman said, "I want a second opinion." "OK," the doctor said. "You're ugly too."

[a] Jokes at the expense of doctors (ie, doctors are the butt of the joke) include jokes about medical physicians or the medical profession, the health care system, and poking fun at advice from doctors or the patient-doctor relationship.

[b] We corrected minor typographical errors and misspellings to improve readability and further de-identify study participants. Emoticon translations: :) smiley face; ;) smiley face with clever wink.

Figure 2. Flow diagram for study inclusion of doctor jokes.



Measures

Characteristics of Jokers

We operationally defined a Lifestreaming Panel participant who posted a doctor joke during the 6-month time period as a joker. We compared sociodemographic characteristics for jokers to nonjokers (participants who did not post a doctor joke on their Facebook wall during the 6-month observation period) and used the total number of Facebook friends at the time of data collection as the size of the participant's social network. Because of the skewed distribution, we report the median number of friends, as well as number of Facebook fan pages and groups to which the participants belonged.

Characteristics of Doctor Jokes

We categorized each joke into 1 of 2 groups: those that were made at the expense of doctors (defined as jokes in which medical doctors or the medical profession, doctors' advice, the doctor-patient relationship, or the health care system were the butt of the joke) and those in which this was not the case. We also determined whether or not the joke was a pun (ie, dependent on multiple definitions of a word), used dirty humor (defined as including foul language, sexual content, racism, degradation of women, reference to human waste, or that were otherwise in poor taste), referred to popular media or culture (television, movies, comedians, or advertising), or related to either politics or current events.

We estimated joke effectiveness in 2 different ways. First, we determined whether the joke received at least one electronic laugh (from a person other than the joker) anywhere in the response from the social network. We operationally defined an electronic laugh as containing one of the following responses: laughing out loud (LOL), rolling on the floor laughing (ROTFL), or an interjection (eg, “baha” or “haha”). For the 11 jokes in which an electronic laugh was included with root post by the joker, it was not considered evidence for (or against) a joke’s success [21].

During the course of the 6-month data collection period, the Facebook feature known as the “like” button that allows members of a network to show enthusiasm for a particular post became popular. As a secondary measure of joke success, we collected the total number of Facebook likes for the 225 jokes in our data that appeared after Harris began collecting data on this feature.

Analyses

We used Stata version 13.0 statistical software (StataCorp LP, College Station, TX, USA) for all analyses. The chi-square test for categorical variables and the *t* test for continuous variables were used to compare sociodemographic characteristics of jokers to nonjokers, except for measures of Facebook activity (ie, social network size and the number of fan and group pages to which a participant belonged), which we compared by using a Mann-Whitney test. Missing values were assigned a category for the comparison of characteristics between jokers and nonjokers for categorical variables.

Because some jokes were repeated in our sample, we used generalized linear mixed-effects models that were fit to allow each joke to have a random intercept. Fixed effects in our models included the characteristics of the joke and the covariates for age, sex, and network size as a categorical variable (ie, small vs medium or large social network). To determine if specific characteristics of jokes predicted the success of a joke, we assumed a binomial distribution for our models in which the dependent variable was whether or not the joke received at least one electronic laugh (yes vs no), and a Poisson distribution for our models in which the total number of likes was the dependent variable.

Results

Characteristics of Jokers

Among the 33,326 Facebook users in the Lifestreaming Panel, 29.99% (9994/33,326) mentioned the word doctor during the 6-month observation period, but only 263 (0.79%) posted a doctor joke. Jokers varied by US region (*P*<.001) and were more heavily represented in the Northeast (the region with the highest density of physicians per capita) [27] and the South (Table 2).

Jokers differed from the typical Lifestreaming Panel participant in several ways: they were less educated (eg, 16.0% of jokers reported having graduate training or an advanced degree whereas 20.38% of nonjokers did, *P*=.02) and they were more likely to be divorced, separated, or widowed (vs married, *P*<.01). Jokers had larger social networks than nonjokers: jokers had a median of 227 (IQR 138-369) Facebook friends compared to 132 (IQR 56-270) among nonjokers (*P*<.001).

Table 2. The characteristics of study participants according to joker status.

Sociodemographic characteristics	Joker status[a]		*P* [b]
	Joker (n=263)	Nonjoker (n=33,063)	
US region, n (%)			<.001
Northeast	55 (20.9)	5911 (17.88)	
Midwest	59 (22.4)	8,575 (25.94)	
South	98 (37.4)	11,249 (34.02)	
West	38 (14.5)	6,826 (20.65)	
Unknown	13 (4.9)	502 (1.52)	
Age, mean (SD)	44.9 (0.7)	44.76 (0.08)	.89
Sex, n (%)			.84
Male	90 (34.2)	11,113 (33.61)	
Female	173 (65.8)	21,950 (66.39)	
Race/ethnicity, n (%)			.16
Hispanic	6 (2.3)	1281 (3.87)	
Non-Hispanic White	216 (82.1)	26,764 (80.95)	
Non-Hispanic Black	11 (4.2)	1657 (5.01)	
Other or multiple races	19 (7.2)	2615 (7.91)	
Unknown	11 (4.2)	746 (2.26)	
Marital status, n (%)			<.01
Married	118 (44.9)	15,268 (46.18)	
Never married	62 (23.6)	8689 (26.28)	
Divorced, separated, or widowed	56 (21.3)	4554 (13.77)	
Unknown	27 (10.3)	4552 (13.77)	
Education, n (%)			.02
High school graduate or less	45 (17.1)	4937 (14.93)	
Some college or associate’s degree	117 (44.5)	13,145 (39.76)	
College degree	56 (21.3)	6732 (20.36)	
Graduate training or advanced degree	42 (16.0)	6739 (20.38)	
Unknown	3 (1.1)	1510 (4.57)	
Self-reported annual earnings (US$), n (%)			.08
<35,000	96 (36.5)	9918 (30.00)	
35,000-74,999	78 (29.7)	11,094 (33.55)	
≥75,000	62 (23.6)	9096 (27.51)	
Unknown	27 (10.3)	2955 (8.94)	
Facebook activity, median (IQR)			
Friends	227 (138-369)	132 (56-270)	<.001
Fan pages	128 (57-260)	48 (16-117)	<.001
Groups	20 (9-36)	7 (2-19)	<.001

[a] Based on data from a 6-month observation period (from December 15, 2010 to June 16, 2011). Joker: posted a doctor joke; nonjoker did not post a doctor joke.

[b] Chi-square test used in comparison of proportions, *t* test used in comparison of means, and Man-Whitney test used for numbers of friends, fan pages, and groups on Facebook.

Doctor Jokes

Of the 156 unique jokes, 112 (71.8%) appeared only once in our data. Of the jokes that were repeated, 2 jokes were particularly popular and were repeated approximately 30 times.

Among the 156 unique doctor jokes, the joke was made at the expense of doctors (and/or the health care system) in approximately half (62/156, 39.7%). In addition, 25.0% (39/156) of jokes relied on dirty humor, 19.9% (31/156) were puns, 14.1% (22/156) pertained to popular culture and media, and 5.8% (9/156) related to current events and politics.

Joke Success

Approximately half of all jokes posted (133/321, 41.4%) received electronic laughter. Jokes made at the expense of doctors were more likely to receive electronic laughter although this did not reach statistical significance (46.5% vs 37.3%, *P*=.09). The marginal trend (OR 1.46, 95% CI 0.94-2.29) toward a higher rate of electronic laughter in response to jokes at doctors' expense was only slightly attenuated after adjusting for age, sex, and social network size (Table 3). The likelihood of generating an electronic laugh was not dependent on other joke characteristics in both our univariate and adjusted models.

The median number of Facebook likes for doctor jokes was 2 (IQR 0-19). Ironically, the joke with the greatest number of Facebook likes (49 total likes from a network of 253 friends) was a "doctor, priest, lawyer" joke in which lawyers were the butt of the joke.

We observed similar associations between the characteristics of jokes and receiving Facebook likes as a measure of joke success (Table 3 vs Table 4). Jokes that were at the expense of doctors received nearly 50% more likes (rate ratio [RR] 1.48, 95% CI 0.96-2.27) in our adjusted model (Table 4). However, jokes that used what we deemed dirty humor were actually less likely to receive Facebook likes, whereas jokes that were based on current events/politics received more than double the amount of Facebook likes (RR 2.36, 95% CI 0.97-5.74; *P*=.06).

Table 3. Univariate and adjusted odds ratios (OR) from mixed models (n=321 jokes) for the association between joke characteristics and elicitation of electronic laughter from social network.

Joke characteristic	Univariate		Adjusted[a]	
	OR (95% CI)	*P*	OR (95% CI)	*P*
At the expense of doctors[b]	1.46 (0.94, 2.29)	.10	1.43 (0.91, 2.26)	.12
Dirty humor	1.22 (0.63, 2.35)	.56	1.11 (0.57, 2.16)	.76
Pun	0.99 (0.57, 1.74)	.98	1.01 (0.58, 1.77)	.97
Based on popular culture	0.82 (0.36, 1.86)	.64	0.79 (0.35, 1.81)	.58
Based on current events/politics	1.85 (0.52, 6.64)	.34	1.73 (0.48, 6.20)	.40

[a] Adjusted for age (continuous, years), sex, and network size (categorical, 0-171 vs 172-293 or ≥294 friends).

[b] Jokes at the expense of doctors (ie, doctors are the butt of the joke) include jokes about medical physicians or the medical profession, the health care system, and poking fun at advice from doctors or the patient-doctor relationship.

Table 4. Univariate and adjusted rate ratios (RR) from mixed models (n=225 jokes) for the association between joke characteristics and total Facebook likes from social network.

Joke characteristic	Univariate		Adjusted[a]	
	RR (95% CI)	*P*	RR (95% CI)	*P*
At the expense of doctors[b]	1.48 (0.96, 2.27)	.08	1.48 (0.96, 2.27)	.08
Dirty humor	0.62 (0.36, 1.08)	.09	0.62 (0.36, 1.09)	.10
Pun	1.13 (0.65, 1.97)	.65	1.15 (0.67, 2.00)	.62
Based on popular culture	0.62 (0.32, 1.18)	.14	0.62 (0.32, 1.20)	.16
Based on current events/politics	2.32 (0.96, 5.62)	.06	2.36 (0.97, 5.74)	.06

[a] Adjusted for age (continuous, years), sex, and network size (categorical, 0-171 vs 172-293 or ≥294 friends).

[b] Jokes at the expense of doctors (ie, doctors are the butt of the joke) include jokes about medical physicians or the medical profession, the health care system, and poking fun at advice from doctors or the patient-doctor relationship.

Discussion

Principal Results

To our knowledge, this is the first study to use actual Facebook conversations to examine doctor-related humor. Overall, we found a low prevalence of doctor jokes on Facebook and relatively few Facebook users posting jokes about doctors (and the health care system in general). Interestingly, those who posted a doctor joke were more likely to be divorced, separated, or widowed, and to have larger social networks (ie, more friends on Facebook). Given the previous findings that people who want to improve their social status are more likely to joke, it may be that divorced, separated, or widowed Facebook users tell doctor jokes to appeal to a potential partner but, of course, we cannot distinguish the reason from these data. Although initially it appeared that poking fun at doctors (as compared to doctor jokes that were not made at the expense of physicians) led to more successful jokes (in generating electronic laughs and the total number of Facebook likes), such findings were not statistically significant. In regards to Facebook likes, jokes based on current events and politics appeared to receive greater response from an individual's social network, whereas dirty humor jokes received fewer likes.

We also observed that although most jokes appeared only once in our data, a few jokes were repeated many times. Based on our qualitative review of these more highly repeated jokes, they differed little from other jokes in our study. Although we are not able to determine the reason, these findings demonstrate how certain ideas can spread rapidly throughout social networks [28].

Comparison With Prior Work

In recent years, medicine has not been immune to the impact of social networking sites, and there is growing interest in social networking sites among physicians and biomedical researchers. Social networking sites may have important applications for studying social interaction and communication related to health and medicine [29,30]. The medical community has largely focused on discussing the ethics of doctor-patient interaction on social networking sites [31-37] and professionalism of younger practitioners' exposure via social networking sites [38-44]. Only more recently has interest emerged in using social networking sites to employ health interventions [45,46] and to identify certain health behaviors [47-49]. To date, there have been few empirical studies in the biomedical literature that examined conversations on social networking sites in nonpatient population groups.

Although our study examined doctor jokes posted on Facebook and does not represent a comprehensive analysis of public opinion of the medical profession and health care, our analyses are among the first to examine actual social networking site conversations [50]. Primary analysis of Facebook conversations could provide researchers the ability to examine certain health behaviors and popular opinion pertaining to health and medicine. Furthermore, analysis of social media conversations on a larger scale could have important uses, such as studying US public opinion regarding national health policy, developing new methods for public health surveillance, and for sociological study to understand social support for illness in virtual settings. However, as our study demonstrates, conversations from social media sites contain a mixture of both relevant and (depending on the use) potentially irrelevant material.

Limitations

Our study has several limitations that must be acknowledged. First, considering that Lifestreaming Panel participants gave permission to have their Facebook data recorded, the potential for selection bias cannot be ruled out. Second, we used the term "doctor(s)" to identify posts pertaining to medical physicians, which may underestimate the total conversations pertaining to physicians, medical practitioners, or the medical profession. Given the casual nature of Facebook, we thought that the term doctor would be used more commonly than a more formal term such as "physician." Finally, we identified potential jokes for analysis based on our definition of what constituted a joke in Facebook posts; others might have defined jokes differently than we did.

Conclusions

Despite the inherent limitations of our research, this study demonstrates the potential of using social networking sites for research on health and medicine. The adoption of social networking has resulted in growing interest in using outlets such as Facebook and Twitter in creative ways. In this study, we demonstrate how actual data from Facebook conversations can be used to study doctor-related humor. In addition to serving as an example, this study highlights some of the practical considerations regarding the analysis of data from social networking sites.

Acknowledgments

We thank Rachel P Davis, MS, and Lynne M Morgan, MPH, for assistance with data coding and Harris Interactive for graciously providing the data for this study. Davis was supported by Award Number 5K01AT006162 from the National Institutes of Health. The views expressed herein do not necessarily represent the official views of the National Institutes of Health, the Department of Veterans Affairs, or the US government. The National Institutes of Health, US Department of Veteran Affairs, and Harris Interactive did not have any role in the design and conduct of the study.

Conflicts of Interest

Haney was employed by Harris Interactive during the study, but Harris Interactive had no say or influence over the study design or analysis. There were no other conflicts of interest to declare.

References

1. Fine GA. Humorous interaction and the social construction of meaning: making sense in a jocular vein. Studies in Symbolic Interaction 1984;5:83-104.
2. Davis MS. What's So Funny?: The Comic Conception of Culture and Society. Chicago: University of Chicago Press; 1993.
3. LaFave L. Humor judgments as a function of reference groups identification classes. In: The Psychology of Humor: Theoretical Perspectives and Empirical Issues. The Psychology of Humor: New York, Academic Press; 1972:195-210.

4. LaFave L, Haddad J, Marshall N. Humour judgements as a function of identification classes. Sociology and Social Research 1973;53:184-194.
5. Francis LE. Laughter, the best mediation: humor as emotion management in interaction. Symbolic Interaction 1994 May;17(2):147-163. [doi: 10.1525/si.1994.17.2.147]
6. Romero E, Pescosolido A. Humor and group effectiveness. Human Relations 2008 Mar 01;61(3):395-418. [doi: 10.1177/0018726708088999]
7. Abel MH. Humor, stress, and coping strategies. Humor: International Journal of Humor Research 2002;15(4):365-381. [doi: 10.1515/humr.15.4.365]
8. White S, Winzelberg A. Laughter and stress. Humor: International Journal of Humor Research 1992;5(4):343-356. [doi: 10.1515/humr.1992.5.4.343]
9. Light D. The medical profession organizational change: From professional dominance to countervailing power. In: Handbook of Medical Sociology. Upper Saddle River, NJ: Prentice Hall; 2000:201-216.
10. Pescosolido BA, Tuch SA, Martin JK. The profession of medicine and the public: examining Americans' changing confidence in physician authority from the beginning of the 'health care crisis' to the era of health care reform. J Health Soc Behav 2001 Mar;42(1):1-16. [Medline: 11357716]
11. Wolinsky, Frederic W. The professional dominance, deprofessionalization, proletarianization, corporatization perspectives: An overview synthesis. In: The Changing Medical Profession: An International Perspective. New York: Oxford University Press; 1993:11-24.
12. McDonald P. The Independent. 2010 Dec 19. Heard the one about the oldest joke in the world? It's a cracker! URL: http://www.independent.co.uk/voices/commentators/paul-mcdonald-heard-the-one-about-the-oldest-joke-in-the-world-its-a-cracker-2164293.html [accessed 2013-12-06] [WebCite Cache ID 6LfjNz2Go]
13. Landon BE, Reschovsky J, Blumenthal D. Changes in career satisfaction among primary care and specialist physicians, 1997-2001. JAMA 2003;289(4):442-449. [Medline: 12533123]
14. Schwartz MD, Durning S, Linzer M, Hauer KE. Changes in medical students' views of internal medicine careers from 1990 to 2007. Arch Intern Med 2011 Apr 25;171(8):744-749. [doi: 10.1001/archinternmed.2011.139] [Medline: 21518941]
15. Boxer D, Cortés-Conde F. From bonding to biting: conversational joking and identity display. Journal of Pragmatics 1997 Mar;27(3):275-294. [doi: 10.1016/S0378-2166(96)00031-8]
16. Freud S, Strachey J, Gay P. Jokes and Their Relation to the Unconscious. New York: Norton; 1960.
17. Sandford S, Eder D. Adolescent humor during peer interaction. Social Psychology Quarterly 1984;47(3):235-243.
18. Schoen C, Osborn R, Doty MM, Bishop M, Peugh J, Murukutla N. Toward higher-performance health systems: adults' health care experiences in seven countries, 2007. Health Aff (Millwood) 2007;26(6):w717-w734 [FREE Full text] [doi: 10.1377/hlthaff.26.6.w717] [Medline: 17978360]
19. Adamle KN, Ludwick R. Humor in hospice care: who, where, and how much? Am J Hosp Palliat Care 2005;22(4):287-290. [Medline: 16082915]
20. Robinson DT, Smith-Lovin L. Getting a laugh: gender, status, and humor in task discussions. Social Forces 2001 Sep 01;80(1):123-158. [doi: 10.1353/sof.2001.0085]
21. Glenn PJ. Initiating shared laughter in multi - party conversations. Western Journal of Speech Communication 1989 Aug;53(2):127-149. [doi: 10.1080/10570318909374296]
22. Norrick NR. Conversational Joking: Humor in Everyday Talk. Bloomington: Indiana University Press; 1993.
23. Facebook Newsroom. 2013. Facebook key facts URL: http://newsroom.fb.com/Key-Facts [accessed 2014-01-26] [WebCite Cache ID 6MvelIMwP]
24. Attardo S. Linguistic Theories of Humor (Humor Research, No. 1). Berlin: Walter de Gruyter; 1994.
25. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990;43(6):551-558. [Medline: 2189948]
26. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543-549. [Medline: 2348207]
27. Goodman DC, Fisher ES, Bronner KK. Hospital and Physician Capacity Update: A Brief Report from the Dartmouth Atlas of Health Care. Hanover, NH: The Dartmouth Institute for Health Policy and Clinical Practice; 2009 Mar 30. URL: http://www.dartmouthatlas.org/downloads/reports/Capacity_Report_2009.pdf [accessed 2014-01-28] [WebCite Cache ID 6MyC5T1Q8]
28. Watts DJ, Dodds PS. Influentials, networks, and public opinion formation. J Consum Res 2007 Dec;34(4):441-458. [doi: 10.1086/518527]
29. Centola D. Social media and the science of health behavior. Circulation 2013 May 28;127(21):2135-2144. [doi: 10.1161/CIRCULATIONAHA.112.101816] [Medline: 23716382]
30. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]

31. Jain SH. Practicing medicine in the age of Facebook. N Engl J Med 2009 Aug 13;361(7):649-651. [doi: 10.1056/NEJMp0901277] [Medline: 19675328]
32. Lacson SM, Bradley C, Arkfeld DG. Facebook medicine. J Rheumatol 2009 Jan;36(1):211. [doi: 10.3899/jrheum.080750] [Medline: 19208543]
33. Grover M. Defining the patient-physician relationship in the era of facebook. Acad Med 2010 Aug;85(8):1262. [doi: 10.1097/ACM.0b013e3181e5c0d2] [Medline: 20671446]
34. Landman MP, Shelton J, Kauffmann RM, Dattilo JB. Guidelines for maintaining a professional compass in the era of social networking. J Surg Educ 2010;67(6):381-386. [doi: 10.1016/j.jsurg.2010.07.006] [Medline: 21156295]
35. Devi S. Facebook friend request from a patient? Lancet 2011 Apr 2;377(9772):1141-1142. [Medline: 21465700]
36. Fenton CH. A few fears felt using Facebook. BMJ 2011;343:d5619. [Medline: 21914737]
37. Osman A, Wardle A. Is it time for medicine to update its Facebook status? BMJ 2011;343:d6334. [Medline: 21971167]
38. Garner J, O'Sullivan H. Facebook and the professional behaviours of undergraduate medical students. Clin Teach 2010 Jun;7(2):112-115. [doi: 10.1111/j.1743-498X.2010.00356.x] [Medline: 21134159]
39. Gray K, Annabell L, Kennedy G. Medical students' use of Facebook to support learning: insights from four case studies. Med Teach 2010;32(12):971-976. [doi: 10.3109/0142159X.2010.497826] [Medline: 21090950]
40. Alkhateeb FM, Clauson KA, Latif DA. Pharmacist use of social media. Int J Pharm Pract 2011 Apr;19(2):140-142. [doi: 10.1111/j.2042-7174.2010.00087.x] [Medline: 21385246]
41. Giordano C, Giordano C. Health professions students' use of social media. J Allied Health 2011;40(2):78-81. [Medline: 21695367]
42. McGinnis MS. Using Facebook as your professional social media presence. Imprint 2011;58(4):36-39. [Medline: 22017023]
43. Moubarak G, Guiot A, Benhamou Y, Benhamou A, Hariri S. Facebook activity of residents and fellows and its impact on the doctor-patient relationship. J Med Ethics 2011 Feb;37(2):101-104. [doi: 10.1136/jme.2010.036293] [Medline: 21160080]
44. Thompson LA, Black E, Duff WP, Paradise Black N, Saliba H, Dawson K. Protected health information on social networking sites: ethical and legal considerations. J Med Internet Res 2011;13(1):e8 [FREE Full text] [doi: 10.2196/jmir.1590] [Medline: 21247862]
45. Moreno MA, Vanderstoep A, Parks MR, Zimmerman FJ, Kurth A, Christakis DA. Reducing at-risk adolescents' display of risk behavior on a social networking web site: a randomized controlled pilot intervention trial. Arch Pediatr Adolesc Med 2009 Jan;163(1):35-41. [doi: 10.1001/archpediatrics.2008.502] [Medline: 19124701]
46. Gold J, Pedrana AE, Stoove MA, Chang S, Howard S, Asselin J, et al. Developing health promotion interventions on social networking sites: recommendations from The FaceSpace Project. J Med Internet Res 2012;14(1):e30 [FREE Full text] [doi: 10.2196/jmir.1875] [Medline: 22374589]
47. Moreno MA, Christakis DA, Egan KG, Brockman LN, Becker T. Associations between displayed alcohol references on Facebook and problem drinking among college students. Arch Pediatr Adolesc Med 2012 Feb;166(2):157-163 [FREE Full text] [doi: 10.1001/archpediatrics.2011.180] [Medline: 21969360]
48. Moreno MA, Christakis DA, Egan KG, Jelenchick LA, Cox E, Young H, et al. A pilot evaluation of associations between displayed depression references on Facebook and self-reported depression using a clinical scale. J Behav Health Serv Res 2012 Jul;39(3):295-304 [FREE Full text] [doi: 10.1007/s11414-011-9258-7] [Medline: 21863354]
49. Moreno MA, Egan KG, Brockman L. Development of a researcher codebook for use in evaluating social networking site profiles. J Adolesc Health 2011 Jul;49(1):29-35 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.04.015] [Medline: 21700153]
50. Martinello N, Donelle L. Online conversations among Ontario university students: environmental concerns. Inform Health Soc Care 2012 Sep;37(3):177-189. [doi: 10.3109/17538157.2012.678448] [Medline: 22713097]

Abbreviations

LOL: laughing out loud
ROTFL: rolling on the floor laughing
RR: rate ratio

Edited by G Eysenbach; submitted 27.06.13; peer-reviewed by A Roundtree, J Yaphe; comments to author 17.09.13; revised version received 10.12.13; accepted 15.01.14; published 13.02.14.

Please cite as:
Davis MA, Haney CS, Weeks WB, Sirovich BE, Anthony DL
Did You Hear the One About the Doctor? An Examination of Doctor Jokes Posted on Facebook
J Med Internet Res 2014;16(2):e41
URL: http://www.jmir.org/2014/2/e41/
doi:10.2196/jmir.2797
PMID:24550095

Review

Are Health Behavior Change Interventions That Use Online Social Networks Effective? A Systematic Review

Carol A Maher[1], PhD; Lucy K Lewis[1], PhD; Katia Ferrar[2], PhD; Simon Marshall[3], PhD; Ilse De Bourdeaudhuij[4], PhD; Corneel Vandelanotte[5], PhD

[1]Health and Use of Time Group, University of South Australia, Adelaide, Australia

[2]Exercise for Health and Human Performance Group, University of South Australia, Adelaide, Australia

[3]School of Medicine, University of California, San Diego, CA, United States

[4]Department of Movement and Sports Sciences, University of Ghent, Ghent, Belgium

[5]Institute for Health and Social Sciences, Central Queensland University, North Rockhampton, Australia

Corresponding Author:
Carol A Maher, PhD
Health and Use of Time Group
University of South Australia
GPO Box 2471
Adelaide, 5001
Australia
Phone: 61 8 8302 2315
Fax: 61 8 8302 2766
Email: carol.maher@unisa.edu.au

Abstract

Background: The dramatic growth of Web 2.0 technologies and online social networks offers immense potential for the delivery of health behavior change campaigns. However, it is currently unclear how online social networks may best be harnessed to achieve health behavior change.

Objective: The intent of the study was to systematically review the current level of evidence regarding the effectiveness of online social network health behavior interventions.

Methods: Eight databases (Scopus, CINAHL, Medline, ProQuest, EMBASE, PsycINFO, Cochrane, Web of Science and Communication & Mass Media Complete) were searched from 2000 to present using a comprehensive search strategy. Study eligibility criteria were based on the PICOS format, where “population” included child or adult populations, including healthy and disease populations; “intervention” involved behavior change interventions targeting key modifiable health behaviors (tobacco and alcohol consumption, dietary intake, physical activity, and sedentary behavior) delivered either wholly or in part using online social networks; “comparator” was either a control group or within subject in the case of pre-post study designs; “outcomes” included health behavior change and closely related variables (such as theorized mediators of health behavior change, eg, self-efficacy); and “study design” included experimental studies reported in full-length peer-reviewed sources. Reports of intervention effectiveness were summarized and effect sizes (Cohen’s *d* and 95% confidence intervals) were calculated wherever possible. Attrition (percentage of people who completed the study), engagement (actual usage), and fidelity (actual usage/intended usage) with the social networking component of the interventions were scrutinized.

Results: A total of 2040 studies were identified from the database searches following removal of duplicates, of which 10 met inclusion criteria. The studies involved a total of 113,988 participants (ranging from n=10 to n=107,907). Interventions included commercial online health social network websites (n=2), research health social network websites (n=3), and multi-component interventions delivered in part via pre-existing popular online social network websites (Facebook n=4 and Twitter n=1). Nine of the 10 included studies reported significant improvements in some aspect of health behavior change or outcomes related to behavior change. Effect sizes for behavior change ranged widely from −0.05 (95% CI 0.45-0.35) to 0.84 (95% CI 0.49-1.19), but in general were small in magnitude and statistically non-significant. Participant attrition ranged from 0-84%. Engagement and fidelity were relatively low, with most studies achieving 5-15% fidelity (with one exception, which achieved 105% fidelity).

Conclusions: To date there is very modest evidence that interventions incorporating online social networks may be effective; however, this field of research is in its infancy. Further research is needed to determine how to maximize retention and engagement, whether behavior change can be sustained in the longer term, and to determine how to exploit online social networks to achieve mass dissemination. Specific recommendations for future research are provided.

(J Med Internet Res 2014;16(2):e40) doi:10.2196/jmir.2952

KEYWORDS

systematic review; social network; behavior change; intervention; Internet; physical activity; weight loss

Introduction

Preventing and minimizing the impact of non-communicable diseases are some of the greatest challenges facing modern society. Key health behaviors, such as physical inactivity, smoking, obesity, poor diets, and alcohol misuse are among the most common causes of disease and premature deaths in both developed countries and, increasingly, developing countries [1-3]. For example, it has been estimated that approximately 2.6 million years of life are lost in England and Wales each year due to preventable disease burden [4]. Cost-effective, mass-reach public health interventions are needed to optimize health and well-being and minimize health care costs of lifestyle diseases.

A variety of media have been used to deliver mass-reach health campaigns, including television, radio, and billboard advertising [5], Web-based interventions [6], and recently, online social networks [7]. Online social networks have seen enormous growth in popularity in recent years and account for approximately one-quarter of all time spent online [8,9]. At present, there are many uncertainties as to whether, and how, online social networks might be harnessed to improve health: for example, how to handle privacy issues and whether people even desire to use online social networks to engage in health behavior change. Certainly, online social networks appear to offer considerable potential for delivery of public health campaigns, for several reasons. First, they can reach very large audiences (eg, Facebook, the world’s largest social networking website, has 1.1 billion users each month) [10]. Second, messages can be delivered via existing contacts, which may be more influential than health messages delivered via traditional marketing strategies [11]. Third, unlike traditional Web-based interventions [6], online social networks typically achieve high levels of user engagement and retention [10]. Finally, social media requires users to actively engage and generate content, which may well be more influential than traditional websites and advertising that are typically more passive in nature [12].

A number of studies have recently attempted to use online social networking strategies to instigate health behavior change. However, despite the large potential for behavior change and the immense popularity of online social networks, it is unclear how effective this approach has been across a range of different population groups and health behaviors. Therefore, this study aimed to systematically review the current level of evidence regarding the effectiveness of online social network health behavior interventions to influence tobacco and alcohol consumption, dietary intake, physical activity, and sedentary behavior.

Methods

Information Sources and Search Strategy

This review was undertaken and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [13].

A preliminary search protocol was drafted and included terms for social media, the Internet, and the relevant health behaviors for this review. The search strategy was reviewed by experts in the area of online interventions (three members of the authorship team—CV, SM, and IDB) and an academic librarian before being finalized [14]. The final search was conducted on December 12, 2012 and included eight electronic databases: Scopus, CINAHL, Medline, ProQuest, EMBASE, PsycINFO, Cochrane, Web of Science and Communication & Mass Media Complete. Each database was searched individually and the search strategy for one database, Medline, is presented in Table 1. The search was limited to the English language, humans, and the year of publication from 2000 to present, and the search terms mapped to MeSH headings wherever possible. The reference lists of included studies and relevant systematic reviews were screened to identify further eligible studies.

Table 1. The search strategy as used in Medline.

Search category	Search terms
1. Social media	Social Networking/[a] OR social network*.mp.[b] OR social media.mp. OR Social Media/ OR (Facebook OR LinkedIn OR Twitter OR Badoo OR Orkut OR Qzone OR Xing OR Tencent OR Weibo OR Mixi OR Sina Weibo OR Hyves OR Skyrock OR Odnoklassniki OR Wer-kennt-wen OR V Kontakte OR Tuenti OR MySpace).mp.
2. Internet	online.mp. OR Internet/ OR internet.mp. OR web.mp.
3. Health behaviors	(cigarette.mp. OR Tobacco/ OR tobacco.mp. OR Smoking/ OR smoking.mp. OR Smoking Cessation/ OR nicotine.mp. OR Nicotine/) OR (alcohol.mp. OR Alcohol Drinking/ OR “binge drink*”.mp. OR “alcohol drink*”.mp.) OR (Motor Activity/ OR “physical activit*”.mp. OR “motor activit*”.mp. OR PA.mp. OR exercise.mp. OR Exercise/ OR exercis*.mp. OR sport*.mp. OR Sports/ OR MVPA.mp. OR Sedentary Lifestyle/ OR sedentar*.mp. OR sitting.mp. OR “screen time”.mp. OR inactiv*.mp. Television/ OR television.mp. OR TV.mp. OR Video games/ OR “video gam*”.mp.) OR (Diet/ OR diet*.mp. OR nutrition*.mp. OR “healthy eating”.mp. OR Food Habits/ OR Fruit/ OR fruit.mp. OR Vegetables/ OR vegetable*.mp. OR “snack food*”.mp. OR snack*.mp. OR” soft drink*”.mp. OR Carbonated beverages/) OR (Health Behavior/ OR “health behav*”.mp.)
4. Combined	1 AND 2 AND 3

[a] “/” denotes MeSH headings

[b] “.mp” denotes keyword search

Study Selection

As per best practice for systematic reviews [15,16], eligibility of studies for inclusion in the review was determined by two independent reviewers (KF and either CM or LL), with results compared and disagreements discussed until consensus was reached. First, search results were screened based on the title and abstract and where eligibility was unclear or the abstract was unavailable the full text was obtained. The eligibility criteria were then applied to the full-text studies to determine inclusion in the review.

Eligibility Criteria

Population

Adults or children were included, regardless of health status (healthy or participants with specific health conditions or diseases).

Intervention

Studies were included that reported an online intervention delivered either wholly or in part, using an online social network to deliver a health behavior change intervention. The online social network intervention could be delivered using an existing online social networking platform (eg, intervention delivered via either a “generic” pre-existing social networking website such as Facebook or Twitter, or a health-specific pre-existing social networking website, such as FatSecret) or a purpose-built intervention website incorporating social networking capabilities. In the case of purpose-built websites, studies had to explicitly describe their website as using social networking to be included. Interventions delivered via purpose-built websites which facilitated a degree of interactivity between participants (eg, a discussion board) but did not specifically describe the intervention as being or involving a “social network” were excluded.

Control or Comparator

Any comparator was acceptable (ie, a traditional control group, an alternative intervention, or a within subject pre-post design).

Outcomes

The online social media intervention had to target one of the following individual modifiable health behaviors identified by the World Health Organization as leading risk factors for global disease burden [3]: tobacco smoking, alcohol use, physical inactivity, or diet. For inclusion in the review, the study had to report data regarding the effectiveness of behavior change (eg, change in physical activity behavior [min/d]). Additionally, studies were included if they reported variables closely related to behavior change; this included potential mediators of behavior change (eg, dietary awareness or physical activity self-efficacy), or “downstream” variables (ie, variables that may have conceivably been impacted by health behavior change; eg, quality of life or body weight).

Study Design

Only experimental studies that were reported in peer-reviewed journals or as peer-reviewed full conference papers were included. Ecological studies, as well as studies employing small samples (eg, case studies), were eligible to be included in the review. Relevant systematic reviews were retained and the reference lists searched for additional relevant studies. Conference abstracts and theses were excluded.

Data Collection Process and Data Items

Data extraction was conducted using a standardized form developed specifically for this review (see Multimedia Appendix 1), based upon that used by Davies et al [6]. For each included study, pairs of reviewers independently extracted data (CV/LL, CM/KF), with disagreements resolved by checking and discussing the original study until consensus was reached. Percent agreement between reviewers for data extraction was 88%, with the main discrepancies relating to classification of target behavior in the case of weight loss studies (whether the intervention targeted diet or weight loss or both). Extracted information included study participants (population, sample size, participation rate, attrition rate, recruitment method, setting), study design and duration of follow-up, behavior targeted, intervention description (including format, intensity, duration, and theoretical basis), and outcome measures used.

Risk of Methodological Bias

The included studies varied widely in terms of research design, making selection of an appropriate risk of bias assessment tool difficult. After extensive discussion among the research team, a tool was devised based upon the CONSORT checklist [17]. The tool comprised 25 items, with items scored as 1 or 0 if the studies satisfactorily met/didn't meet the criteria (see Multimedia Appendix 2), with a higher score indicating lower risk of methodological bias. While intended for controlled trials, the team felt that the majority of items (20 out of 25) were applicable to other study designs and that the weaker study designs rightly should receive a lower score than studies utilizing a controlled trial design. Each study was also ranked using the 2011 Centre for Evidence Based Medicine Levels of Evidence, where Level 1 signifies systematic review of randomized trials; Level 2 randomized trial or observational study with dramatic effect; Level 3 non-randomized controlled cohort/follow-up study; Level 4 case-series, case-control studies, or historically controlled studies; and Level 5 mechanism-based reasoning [18]. Rating of studies was conducted independently by pairs of reviewers (CM/LL, CV/KF) with any differences resolved by discussion. Percent agreement between reviewers for the scoring of risk of methodological bias was 81%, with the most common points of discrepancy relating to whether the trial was registered and whether a trial protocol was accessible.

Summary Measures and Synthesis of Results

The primary outcome measure was health behavior change (for example, physical activity and dietary behaviors). Secondary outcome measures related to behavior change were also examined. These could be either "downstream" from behavior change (ie, outcomes brought about by sustained behavior change, for example, change in body weight) or "upstream" (ie, theorized mediators of behavior change, such as knowledge, attitudes, or self-efficacy).

To determine whether the interventions had a significant impact on behavior, we evaluated and coded individual outcomes. In studies without a control group, a positive outcome was recorded if there was a statistically significant change across time. In the case of controlled trials, studies were coded as having a positive outcome if statistically significant differences between groups across time were reported. In the case of controlled trials where the intervention was compared with an alternative intervention (as opposed to a no-intervention control group) and there was a significant improvement in both groups, but not between groups, this was coded as a "suggested positive" outcome. To allow for comparison across studies, effect sizes (Cohen's d and 95% confidence intervals) were calculated according to formulas published by Lipsey and Wilson [19] (online calculators available at [20]). The magnitude of the effect sizes were classified using descriptors proposed by Thalheimer and Cook, where effect sizes ≥ -0.15 and <0.15 are "negligible", ≥ 0.15 and <0.40 are "small", ≥ 0.40 and <0.75 are "medium", ≥ 0.75 and <1.10 are "large", ≥ 1.10 and <1.45 are "very large", and ≥ 1.45 are "huge" [21].

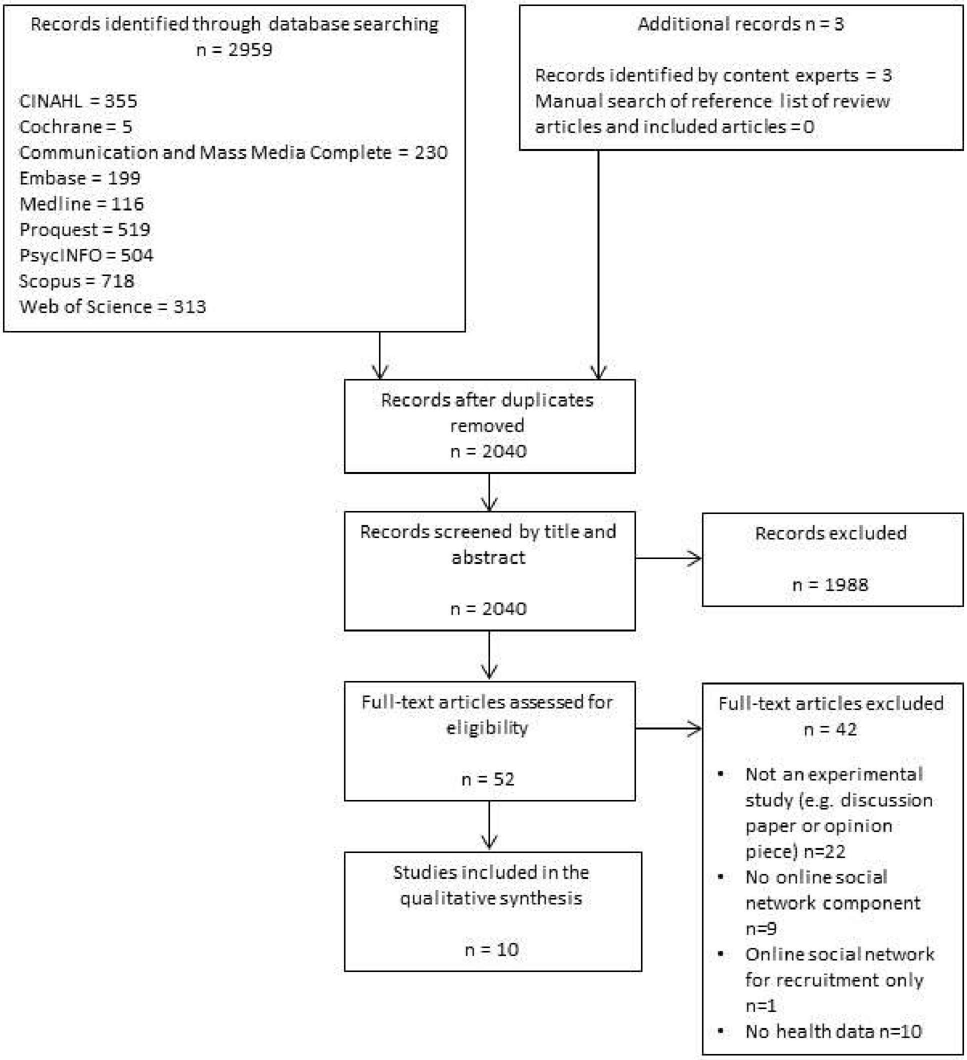
Attrition, engagement, and fidelity with the social networking component of the interventions were scrutinized where sufficient data were presented to allow this. Fidelity was calculated by comparing the actual engagement with the intended dosage.

Results

Study Selection

A total of 2040 studies were identified from the database search following removal of duplicates. The flow of studies through the review is summarized in Figure 1. Ten articles reporting data regarding the effectiveness of online social networking behavior change interventions were included in the review.

Figure 1. Flow of studies through the review.



Study Characteristics

A summary of the key characteristics of included studies is presented in Multimedia Appendix 3. The total number of participants across the 10 studies was 113,988. The studies typically reported high rates of female participation: on average 83.3% of participants were female. The targeted health behaviors were diet/weight loss (n=2), physical activity (n=3), or a combination of diet/weight loss and physical activity (n=5). No eligible studies targeted smoking or alcohol consumption. Five studies involved interventions delivered solely via online social networks [22-26] and the remaining studies involved interventions that used online social networks in conjunction with other intervention strategies [27-31], including standalone intervention websites, printed materials, and provision of supplemental equipment such as kitchen scales and utensils. Only three interventions were reported to be theoretically-based; of these, Social Cognitive Theory was reported in two studies [27,28], and Social Learning Theory [24] was reported in a single study. One further study used a behavior change theory during evaluation (the Theory of Planned Behavior), but not during intervention development [22]. All of the interventions facilitated or encouraged daily use. Interventions ranged in duration from 5 days to 6 months. No studies reported a long-term follow-up to determine whether outcomes were sustained beyond the intervention period.

Study Methodology

Three key study types were identified: (1) large-scale evaluations of “live” interventions, typically with >1000 participants (four studies with sample sizes ranging from 545 to 107,907) [22,24-26], (2) medium-scale, tightly-controlled randomized controlled trials, typically with approximately 100 participants (four studies with sample sizes ranging from 52 to 134) [27-29,31], and (3) small pilot studies, each with 10 participants (two studies) [23,30]. In all, five studies were randomized controlled trials (RCT) [22,27-29,31], one was a randomized cross-over study [23], and four were single group pre-post studies [24-26,30]. Of the six studies that utilized a separate control group or arm (crossover study), only one had a “true” (ie, no-intervention) control [29], with the others comparing the online social networking intervention with an alternative intervention (in five cases the alternative intervention was Web-based [22,23,27,28,31], and in three cases the alternative intervention involved an online social networking component [22,23,27].

Recruitment Methods and Rates

Eight of the 10 included studies recruited participants to their interventions (the other two were evaluations based upon existing users of commercial online social networks [25,26]). Of these, six described their recruitment strategies, with all of these studies reporting a variety of traditional recruitment methods, such as advertising with flyers [27,29,31], mainstream media [22,28], and mass emails [27-29,31]. Only one study [27] reported using an online social media campaign, in addition to other recruitment methods. Participation rates varied widely, ranging from 33% [22] to 89% [27]. However, it is important to note that participation rates were only reported relative to the total number of volunteers coming forward and not the total number of people exposed to recruitment materials.

Risk of Methodological Bias

Risk of methodological bias scores ranged from high (19.5 out of 25) [28], to low (0.5 out of 25) [26]; for full details, see Multimedia Appendix 2. In general, the large-scale “live” interventions scored poorly on the risk of bias assessment (range 0.5-4 out of 25, with the exception of Brindal et al [22], which scored 16.5 out of 25), the medium-scale RCTs scored highest (range 11-19.5 out of 25), and the pilot studies scored poorly (range 4-8.5 out of 25).

Most studies met the CONSORT requirements to provide a strong scientific rationale and described their interventions clearly. However, none of the studies met the stringent guidelines for quality reporting of trial results, which requires provision of effect size estimates and their precision. Only one study reported that participants were blinded to the treatment condition [22]. Attrition rates were reported by eight studies, while participation rates were reported in only five studies.

Intervention and Follow-Up Duration

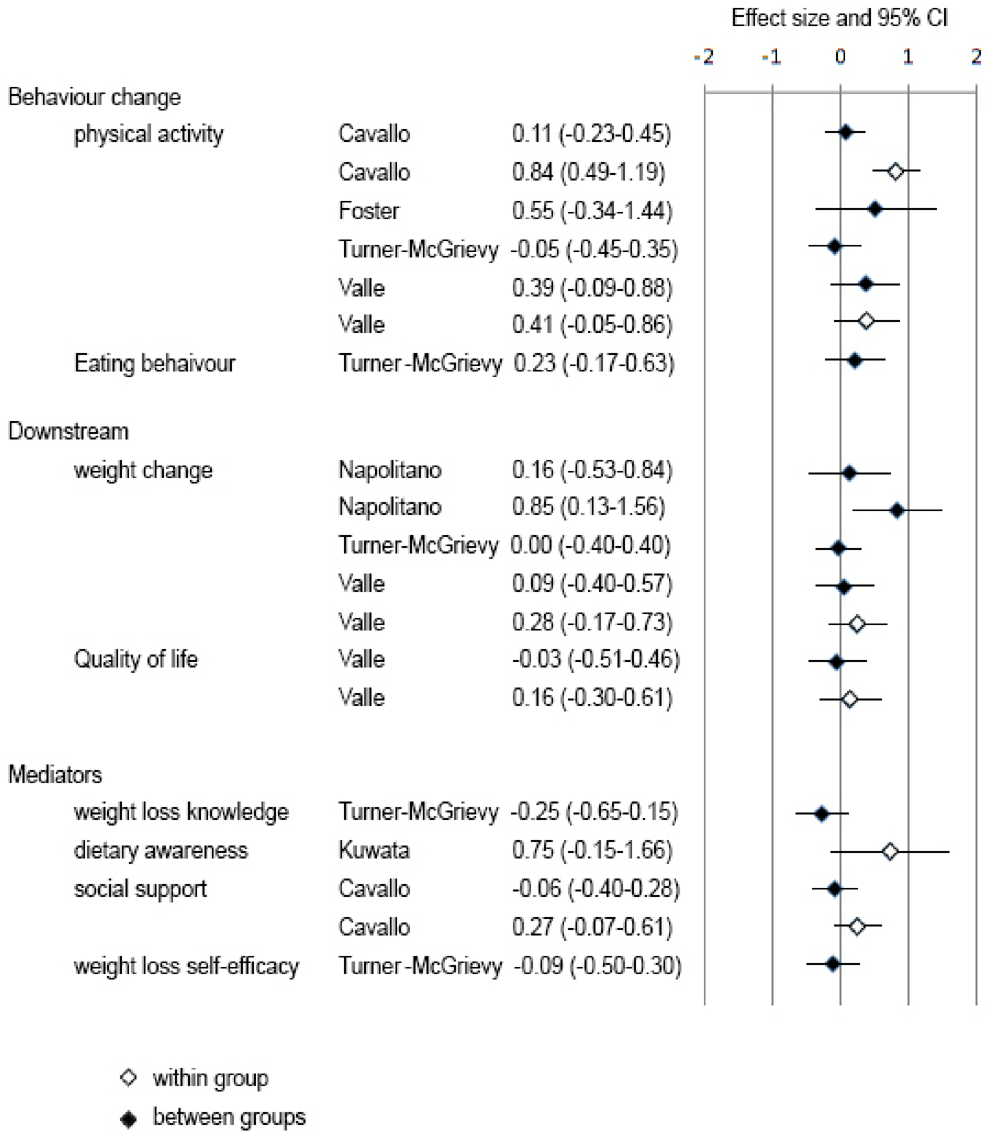
Interventions ranged from 5 days [23] to 6 months in duration [28]. No studies reported follow-up of outcomes and maintenance of behavior change beyond the end of the intervention itself.

Efficacy

Four studies (three pre-post studies and one cross-over study) reported significant improvement in an outcome measure, namely weight loss (n=2) [25,26], physical activity (n=1) [23], and dietary awareness (n=1) [30]. A further four studies, all randomized controlled trials employing alternative intervention controls, reported evidence suggestive of improvement (ie, both groups improved significantly over time, though there was no significant difference between groups) [22,27,28,31]. The remaining two studies [24,29] reported mixed findings; for example, Napolitano et al’s randomized controlled trial reported significant weight loss in the Facebook Plus group relative to controls over time, but not the Facebook group relative to controls [29] (Multimedia Appendix 3).

Meta-analysis was not completed due to the relatively small number of studies included and the wide variety of interventions, comparators, and study designs employed in the studies. However, effect sizes were calculated for the six studies that provided sufficient data to do so (Figure 2). One RCT [29] employed a “true” (ie, no intervention) control; therefore, between group differences are presented. The other three RCTs providing sufficient data to calculate effect sizes used alternative interventions as their control condition [27,28,31]. Therefore, wherever possible, both between group and within group effect sizes (based on pre and post data for the intervention group only) were calculated for these studies. Despite numerous studies reporting statistically significant changes (refer to Table 2), few studies returned significant effects when the 95% confidence intervals were calculated. The magnitude of the effect sizes are summarized below.

Figure 2. Forest plot of effect sizes for behavior change, downstream, and mediator variables.



			Effect size and 95% CI	
Behaviour change				
	physical activity	Cavallo	0.11 (-0.23-0.45)	◆
		Cavallo	0.84 (0.49-1.19)	◇
		Foster	0.55 (-0.34-1.44)	◆
		Turner-McGrievy	-0.05 (-0.45-0.35)	◆
		Valle	0.39 (-0.09-0.88)	◆
		Valle	0.41 (-0.05-0.86)	◇
	Eating behaivour	Turner-McGrievy	0.23 (-0.17-0.63)	◆
Downstream				
	weight change	Napolitano	0.16 (-0.53-0.84)	◆
		Napolitano	0.85 (0.13-1.56)	◆
		Turner-McGrievy	0.00 (-0.40-0.40)	◆
		Valle	0.09 (-0.40-0.57)	◆
		Valle	0.28 (-0.17-0.73)	◇
	Quality of life	Valle	-0.03 (-0.51-0.46)	◆
		Valle	0.16 (-0.30-0.61)	◇
Mediators				
	weight loss knowledge	Turner-McGrievy	-0.25 (-0.65-0.15)	◆
	dietary awareness	Kuwata	0.75 (-0.15-1.66)	◇
	social support	Cavallo	-0.06 (-0.40-0.28)	◆
		Cavallo	0.27 (-0.07-0.61)	◇
	weight loss self-efficacy	Turner-McGrievy	-0.09 (-0.50-0.30)	◆

Table 2. Summary of intervention effects on behavior, downstream, and mediator outcome measures.[a]

Study	Behavior outcomes			Downstream outcomes		Mediators						
	PA[b]	Energy intake	Eating behavior	Weight loss	QOL[c]	Weight loss knowledge	Dietary awareness	PA awareness	Health attitudes	PA self-efficacy	Weight self-efficacy	Social support
Brindal et al [22]				+								
Cavallo et al [31]	+											±
Foster et al [23]	++											
Freyne et al [24]									±*			
Kuwata et al [30]							++	−				
Ma et al [25]				++								
Napolitano et al [29]	−			±*						−	−	−
Sugano & Yamazaki [26]				++								
Turner-McGrievy & Tate [28]	(+)	(+)	(+)	(+)		(+)					(+)	
Valle et al [27]	+			+	±*							

[a] +: within-group significant improvements, in RCT with alternative intervention; ±: within-group mixed results; some significant improvements, some no change, in RCT with alternative intervention; ++: significant improvement; −: no significant change; ±*: mixed results; some subscales showed significant improvement, some showed no significant change; (+) within-group improvements but significance not reported, in RCT with alternative intervention.

[b] PA: physical activity

[c] QOL: quality of life

Behavior Change

Of the four studies that investigated physical activity behavior change and reported sufficient data to enable effect size calculation, the effect size in one study was classified as negligible [28], two as medium (between groups) [23,27], and one as large (between groups) [31]. A small effect size was observed for the one study measuring eating behavior [28].

Downstream Variables

Three studies measured weight change and reported sufficient data to permit effect size calculation. Effect sizes ranged from negligible [28], to small [27], to large [29] (Facebook Plus versus control). One study measured quality of life and reported negligible to small effects [27].

Mediators

A small negative effect size was calculated for the single study that measured weight loss knowledge, though it is worth noting this study employed a substantial alternative intervention control [28]. The same study showed a negligible effect for weight loss self-efficacy. A single study found a negligible to small effect for social support [31], and one study showed a large effect for dietary awareness [30].

Attrition, Engagement, and Fidelity

Attrition rates (ie, participant dropout over the course of the study) varied by study design, with the small scale pilot studies reporting the lowest attrition (0%), the mid-sized RCTs reporting low attrition (4 [29] -23% [27]) and the large live trials reporting high attrition (ranging from 41% in [25] to 84% in [22]; note that attrition rates were not reported in [24] and [26]). Where possible, we examined engagement with the social networking component of the intervention in each study and compared it with the intended dosage, to provide an indication of fidelity. Results (Table 3) showed that fidelity was generally quite low. With the exception of Foster et al [23], which achieved higher usage rates than intended (105%), the other studies reporting usage rates were only 5 to 15% of that intended.

All three studies that reported how engagement changed over the course of the intervention found that it gradually declined [24,27,28]. Studies that compared a social intervention with a non-social control reported that the social intervention achieved higher engagement [22-24] and had higher user satisfaction [22] than the non-social intervention. Weight loss was significantly associated with engagement [22,26] in the two studies that reported this subgroup analysis. Sugano and Yamazaki [26] also reported that extent of social interaction was positively associated with weight loss.

Table 3. Summary of engagement and fidelity with the interventions.

Study	Intended number of uses	Usage rates	
		Actual usage (=“engagement”)	% intended: actual (=“fidelity”)
Brindal et al [22]	84	6.0	7.1%
Cavallo et al [31]	84	5.1	6.1%
Foster et al [23]	21	22	104.8%
Ma et al [25]	133	6.1	4.6%
Turner-McGrievy & Tate [28]	364	54.6	15.0%
Valle et al [27]	84	4.6	5.5%

Discussion

Principal Findings

This systematic review found modest evidence that online social network interventions may be effective, with 9 of the 10 included studies reporting significant improvements in some aspect of health behavior or related outcomes. However, effect sizes for behavior change were generally small.

This review identified that online social network-based interventions to date have taken one of two key approaches: (1) some have developed interventions that used popular existing online social networking websites, such as Facebook and Twitter, and (2) others have developed standalone, health-focused online social networks. Results suggest that standalone health-focused online social networks can be effective for the users they retain over a period of time; however, poor retention is an issue, with roughly 50% or more of users who sign up failing to stay in the intervention for its duration, and for those who do, engagement is generally low. It may also be argued that a drawback of health-focused online social networks is that they are likely to attract motivated individuals who were already contemplating changing their health behavior.

Using popular existing social network sites may address issues of reach, engagement, and retention. For example, Facebook reports that 61% of its total users log in daily [32]. Certainly, the studies included in this review [23,27,29,31] that used Facebook managed to retain a high proportion of participants across the study period (77-96% of users). However, they typically did not achieve high engagement (5-15%) [27,29,31], with the exception of the Foster study, which achieved 105% of intended use [23]. That engagement was typically low is concerning, given that these studies utilized extensive participant contact, prompting, and email, which are likely to have inflated engagement compared to what might be seen in a more ecologically valid setting. The intervention approach used in the Foster study [23] was considerably different to that used in the other Facebook studies [27,29,31], which might explain the different levels of engagement observed. Foster [23] recruited participants who already knew each other and created a friendly competitive environment with a tally board. In contrast, the other studies have tended to use Facebook and Twitter as a social support tool, where intervention participants (who were strangers to each other) were encouraged to share information and advice [27,29,31]. It could be argued that the approach used by Foster [23] was more in tune with how people use online social networks, given that people more commonly use Facebook to interact with people with whom they share an offline connection as well, rather than using Facebook to interact with new people [33]. Furthermore, entertainment is recognized as being a key motivator for Facebook use [34]; the friendly-competitive tone of the Foster intervention was probably consistent with this. The Foster study only ran for 21 days, considerably less than the other interventions, which each lasted 12 weeks. It seems unlikely that the very high engagement achieved by Foster et al [23] would have been sustained over a longer duration. Despite this, the high engagement achieved in those 3 weeks suggests this friendly-competitive intervention may be a promising approach.

Strengths and Limitations

Strengths of this systematic review are that it was conducted and reported according to PRISMA guidelines [13]. It utilized a rigorous and comprehensive search strategy. Study selection, data extraction, and critical appraisal were completed in duplicate by two members of the research team independently, ensuring the accuracy of the review data.

A key limitation of the review was the heterogeneity of the identified studies. Studies varied in terms of target population, intervention, and study design. Furthermore, only a relatively small number of eligible studies were identified. A large number of academic databases were searched (eight), and an academic librarian was consulted regarding which databases should be used; however, it is always possible that other databases may have uncovered additional studies. These factors limited our ability to synthesize data and reach definitive conclusions. It is also important to note that the included studies varied widely in terms of risk of bias, with some studies scoring very poorly, which reduces the trust that can be placed in their findings. Finally, the possibility of publication bias should also be acknowledged. As with all systematic reviews examining the efficacy of interventions, there is a possibility that studies with null findings have not been published [16], and that the synthesis of data presented here gives an overly favorable account of effectiveness.

Future Research

This review offers preliminary evidence that social networking-based health interventions may be effective in changing behavior. However, this field of research is in its infancy and many questions remain unanswered. It is currently unclear whether social networking-based interventions are equally useful for all health behaviors or whether they may be more effective for some than others. The identified studies only followed participants for a relatively short period (the longest was 6 months). Given that many of the health benefits of health behavior are achieved over a long-term period, further work is needed to examine whether the short-term behavior change achieved in the included studies can be sustained over a longer period, such as 12 months and beyond [35]. It will also be important to determine whether sustained interaction with the user interface is required to sustain behavior change or whether behavior change may persist after interaction with intervention materials ceases. Delivery of interventions that use existing online social networks such as Facebook and Twitter appear to offer particular promise for sustained engagement, due to their high level of user retention and engagement, whether retention and engagement with specific aspects of these platforms (such as a specific app or Facebook group delivering a health intervention) matches this is currently unclear. Innovative approaches reflecting the way people use online social networks (with existing friends and for entertainment) are warranted. In particular, gamification (ie, the use of video gaming elements such as collecting virtual points or badges) in non-gaming situations is an emerging trend in online campaigns and offers promise for improving user experience and engagement [36].

Interestingly, to date, the interventions that have used existing popular online social networks have still used traditional methods of recruitment (eg, flyers, media advertising), and have also been highly controlled (eg, group membership has been closed in order to prevent contamination between study groups). This contrasts with the touted benefits of using online social networks for health intervention, such as the ability to recruit participants via social networks [37], and to virally disseminate interventions on a mass scale [38]. There is clearly a role for tightly controlled randomized controlled trials in order to establish efficacy of an intervention approach; however, ecological study designs, which closely mimic real conditions of social network use, are also required in order to learn how to best exploit viral properties of online social networks for mass dissemination [7]. Cross-disciplinary research pairing health behavior change experts with social marketers may help determine how to most effectively use online social networks for recruitment and mass dissemination.

Recommendations for Future Studies

More studies are needed that attempt to intervene in health behavior using online social networks. The following recommendations for future research may be useful for both health researchers and human-computer interaction researchers who design and implement technology-based interventions integrating social networking to facilitate health interventions:

1. Design social-networking interventions that can be delivered primarily within the social network setting. Provision of a small degree of supplementary equipment or printed resources is reasonable, but it is important to recognize that interventions incorporating multiple physical resources have limited ecological validity.
2. Examine interventions delivered via existing popular social network websites, such as Facebook, given their proven ability to attract and retain participants and potential for mass dissemination. Such interventions should be responsive to the way people use online social networks (predominantly with existing friends and for entertainment).
3. Utilize large sample sizes to ensure they are sufficiently powered to detect effects, should they exist.
4. Involve high quality research methods, such as carefully designed randomized controlled trials.
5. In addition to high-quality efficacy studies, ecologically valid studies such as pragmatic randomized trials are also required to determine interventions' abilities to mass disseminate in a real-world setting.
6. Emphasize online recruitment strategies.
7. Involve long-term follow-up (eg, behavior change at 12 months and beyond).

Conclusion

In conclusion, research using online social networks to bring about health behavior change is still in its early stages of development and, while several studies show promise, much is still to be learned about optimizing these interventions to increase their efficacy. In particular, research is needed to determine how to maximize retention and engagement, whether behavior change can be sustained in the longer term, and to determine how to exploit online social networks to achieve mass dissemination.

Acknowledgments

This study was undertaken with assistance of the University of South Australia Fellowship Support funding scheme.

CM is the recipient of an Australian Research Council Australian Postdoctoral Award Fellowship.

Authors' Contributions

This review was conceived and led by CM. CV, SM, and IDB are internationally recognized experts in the field of online behavior change interventions and were invited to join the authorship team from the outset of the systematic review. All members of the research team have prior experience in conducting systematic reviews. All authors contributed to refining of the search strategy and selection of the search terms. The database searches were undertaken by KF. CM, KF, and LL determined study eligibility, with CV, SM, and IDB shown the results and asked as experts in the field to suggest additional studies. Data extraction and risk of bias scoring was undertaken in duplicate by CM, KF, LL, and CV. All authors contributed to interpretation of results and drafting of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data extraction form.

[PDF File (Adobe PDF File), 42KB - jmir_v16i2e40_app1.pdf]

Multimedia Appendix 2

The Risk of Bias tool, based upon the CONSORT checklist.

[PDF File (Adobe PDF File), 53KB - jmir_v16i2e40_app2.pdf]

Multimedia Appendix 3

Overview of the characteristics and outcomes of the included studies.

[PDF File (Adobe PDF File), 243KB - jmir_v16i2e40_app3.pdf]

References

1. Scarborough P, Bhatnagar P, Wickramasinghe KK, Allender S, Foster C, Rayner M. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs. J Public Health (Oxf) 2011;33(4):527-535 [FREE Full text] [doi: 10.1093/pubmed/fdr033] [Medline: 21562029]
2. Narayan KM, Ali MK, Koplan JP. Global noncommunicable diseases--where worlds meet. N Engl J Med 2010;363(13):1196-1198. [doi: 10.1056/NEJMp1002024] [Medline: 20860499]
3. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380(9859):2224-2260. [doi: 10.1016/S0140-6736(12)61766-8] [Medline: 23245609]
4. Office for National Statistics. 2009. Mortality statistics: Deaths registered in England and Wales (Series DR), 2008 URL: http://www.ons.gov.uk/ons/rel/vsob1/mortality-statistics--deaths-registered-in-england-and-wales--series-dr-/2008/index.html [WebCite Cache ID 6JOp6Egbo]
5. Noar SM. A 10-year retrospective of research in health mass media campaigns: where do we go from here? J Health Commun 2006;11(1):21-42. [doi: 10.1080/10810730500461059] [Medline: 16546917]
6. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of Internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9(1):52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
7. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
8. Habeshian V. Marketing Profs. 2013. Social takes up 27% of time spent online URL: http://www.marketingprofs.com/charts/2013/10582/social-takes-up-27-of-time-spent-online [WebCite Cache ID 6JOn52rFb]
9. comScore Data Mine. 2013. A digital month in Germany: what are consumers doing online? URL: http://www.comscoredatamine.com/2013/08/a-digital-month-in-germany-what-are-consumers-doing-online/ [WebCite Cache ID 6JOl18eZ0]
10. Constine J. Techcrunch. 2013. Facebook’s Q2: Monthly users up 21% YOY to 1.15B, dailies up 27% to 699M, mobile monthlies up 51% to 819M URL: http://techcrunch.com/2013/07/24/facebook-growth-2/ [WebCite Cache ID 6JOmhu1I6]
11. De Bruyn A, Lilien GL. A multi-stage model of word-of-mouth influence through viral marketing. Int J Market Res 2008;25(3):151-163. [doi: 10.1016/j.ijresmar.2008.03.004]
12. Thackeray R, Neiger BL, Hanson CL, McKenzie JF. Enhancing promotional strategies within social marketing programs: use of Web 2.0 social media. Health Promot Pract 2008;9(4):338-343. [doi: 10.1177/1524839908325335] [Medline: 18936268]
13. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

14. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. J Clin Epidemiol 2009;62(9):944-952. [doi: 10.1016/j.jclinepi.2008.10.012] [Medline: 19230612]
15. Higgins JPT, Green S. The Cochrane Collaboration. 2011. Cochrane handbook for systematic reviews of interventions version 5.1.0 URL: http://www.cochrane.org/training/cochrane-handbook [WebCite Cache ID 6Lj4qJejo]
16. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [Medline: 19622552]
17. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. Open Med 2010;4(1):e60-e68 [FREE Full text] [Medline: 21686296]
18. Oxford Centre for Evidence-Based Medicine. 2013. The Oxford levels of evidence 2 URL: http://www.cebm.net/index.aspx?o=5653 [WebCite Cache ID 6JOonrB4j]
19. Lipsey MW, Wilson D. Practical meta-analysis. Thousand Oaks, California: SAGE Publications; 2001.
20. Wilson DB. Practical Meta-Analysis Effect Size Calculator. -. URL: http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD1.php [accessed 2014-01-23] [WebCite Cache ID 6MqcuP70x]
21. Thalheimer W, Cook S. How to calculate effect sizes from published research: a simplified methodology. 2002. URL: http://www.bwgriffin.com/gsu/courses/edur9131/content/Effect_Sizes_pdf5.pdf [WebCite Cache ID 6NAyk1pZ3]
22. Brindal E, Freyne J, Saunders I, Berkovsky S, Smith G, Noakes M. Features predicting weight loss in overweight or obese participants in a web-based intervention: randomized trial. J Med Internet Res 2012;14(6):e173 [FREE Full text] [doi: 10.2196/jmir.2156] [Medline: 23234759]
23. Foster D, Linehan C, Kirman B, Lawson S, James G. Motivating physical activity at work: using persuasive social media for competitive step counting. 2010 Presented at: 14th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek; Oct 6-8, 2010; Tampere p. 111-116. [doi: 10.1145/1930488.1930510]
24. Freyne J, Berkovsky S, Kimani S, Baghaei N, Brindal E. Improving health information access through social networking. 2010 Presented at: 23rd IEEE International Symposium on Computer-Based Medical Systems, CBMS; Oct 12-15, 2010; Perth p. 334-339. [doi: 10.1109/CBMS.2010.6042666]
25. Ma X, Chen G, Xiao J. Analysis of an online health social network. 2010 Presented at: 1st ACM International Health Informatics Symposium, IHI'10; Nov 11-12, 2010; Arlington p. 297-306. [doi: 10.1145/1882992.1883035]
26. Sugano M, Yamazaki C. Behavioral analysis of SNS users with regard to diet. 2011 Presented at: IADIS International Conferences - Web Based Communities and Social Media 2011, Social Media 2011, Internet Applications and Research 2011, Part of the IADIS Multi Conference on Computer Science and Information Systems 2011, MCCSIS; July 20-26, 2011; Rome p. 167-170 URL: http://www.researchgate.net/publication/234001361_Behavioral_Analysis_of_SNS_Users_with_Regard_to_Diet
27. Valle CG, Tate DF, Mayer DK, Allicock M, Cai J. A randomized trial of a Facebook-based physical activity intervention for young adult cancer survivors. J Cancer Surviv 2013;7(3):355-368. [doi: 10.1007/s11764-013-0279-5] [Medline: 23532799]
28. Turner-McGrievy G, Tate D. Tweets, apps, and pods: results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]
29. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013;21(1):25-31. [doi: 10.1002/oby.20232] [Medline: 23505165]
30. Kuwata S, Taniguchi S, Kato A, Inoue K, Yamamoto N, Ohkura T, et al. Metaboli-Net: online groupware system providing counseling guidance for patients with metabolic syndrome. Stud Health Technol Inform 2010;156:65-70. [Medline: 20543340]
31. Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012;43(5):527-532. [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]
32. Kaushal N. Facebook to roll-out country wise metrics of monthly active users & daily active users! PageTraffic Buzz. 2013. URL: http://www.pagetrafficbuzz.com/facebook-rollout-country-wise-metrics-monthly-active-users-daily-active-users/17486/ [accessed 2013-09-05] [WebCite Cache ID 6JOnZtJ4D]
33. Ellison NB, Steinfield C, Lampe C. The benefits of Facebook “friends": social capital and college students’ use of online social network sites. J Comput-Mediat Comm 2007;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
34. Cheung CM, Chiu P, Lee MK. Online social networks: why do students use Facebook? Comput Hum Behav 2011;27(4):1337-1343. [doi: 10.1016/j.chb.2010.07.028]
35. Jepson RG, Harris FM, Platt S, Tannahill C. The effectiveness of interventions to change six health behaviours: a review of reviews. BMC Public Health 2010;10(1):538 [FREE Full text] [doi: 10.1186/1471-2458-10-538] [Medline: 20825660]
36. Deterding S, Sicart M, Nacke L, O'Hara K, Dixon D. Gamification. Using game-design elements in non-gaming contexts. 2011 Presented at: CHI '11 Extended Abstracts on Human Factors in Computing Systems; May 7-12, 2011; Vancouver p. 2425-2428. [doi: 10.1145/1979742.1979575]

37. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
38. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010;376(9748):1261-1271. [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]

Abbreviations

PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QOL: quality of life
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 13.09.13; peer-reviewed by E Brindal, D Foster, S Langrial; comments to author 30.10.13; revised version received 09.12.13; accepted 21.12.13; published 14.02.14.

Please cite as:
Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C
Are Health Behavior Change Interventions That Use Online Social Networks Effective? A Systematic Review
J Med Internet Res 2014;16(2):e40
URL: http://www.jmir.org/2014/2/e40/
doi:10.2196/jmir.2952
PMID:24550083

Review

Social Media and Rating Sites as Tools to Understanding Quality of Care: A Scoping Review

Lise M Verhoef[1], MSc; Tom H Van de Belt[2], PhD; Lucien JLPG Engelen[2,3]; Lisette Schoonhoven[1,2,4], RN, PhD; Rudolf B Kool[1], MD, PhD

[1]IQ healthcare, Radboud University Medical Center, Nijmegen, Netherlands
[2]Radboud REshape & Innovation Center, Radboud University Medical Center, Nijmegen, Netherlands
[3]Emergency Healthcare Network, Radboud University Medical Center, Nijmegen, Netherlands
[4]Faculty of Health Sciences, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Lise M Verhoef, MSc
IQ healthcare
Radboud University Medical Center
Route 114
Geert Grooteplein-Noord 21
Nijmegen, 6525 EZ
Netherlands
Phone: 31 243667308
Fax: 31 243540166
Email: Lise.Verhoef@radboudumc.nl

Abstract

Background: Insight into the quality of health care is important for any stakeholder including patients, professionals, and governments. In light of a patient-centered approach, it is essential to assess the quality of health care from a patient's perspective, which is commonly done with surveys or focus groups. Unfortunately, these "traditional" methods have significant limitations that include social desirability bias, a time lag between experience and measurement, and difficulty reaching large groups of people. Information on social media could be of value to overcoming these limitations, since these new media are easy to use and are used by the majority of the population. Furthermore, an increasing number of people share health care experiences online or rate the quality of their health care provider on physician rating sites. The question is whether this information is relevant to determining or predicting the quality of health care.

Objective: The goal of our research was to systematically analyze the relation between information shared on social media and quality of care.

Methods: We performed a scoping review with the following goals: (1) to map the literature on the association between social media and quality of care, (2) to identify different mechanisms of this relationship, and (3) to determine a more detailed agenda for this relatively new research area. A recognized scoping review methodology was used. We developed a search strategy based on four themes: social media, patient experience, quality, and health care. Four online scientific databases were searched, articles were screened, and data extracted. Results related to the research question were described and categorized according to type of social media. Furthermore, national and international stakeholders were consulted throughout the study, to discuss and interpret results.

Results: Twenty-nine articles were included, of which 21 were concerned with health care rating sites. Several studies indicate a relationship between information on social media and quality of health care. However, some drawbacks exist, especially regarding the use of rating sites. For example, since rating is anonymous, rating values are not risk adjusted and therefore vulnerable to fraud. Also, ratings are often based on only a few reviews and are predominantly positive. Furthermore, people providing feedback on health care via social media are presumably not always representative for the patient population.

Conclusions: Social media and particularly rating sites are an interesting new source of information about quality of care from the patient's perspective. This new source should be used to complement traditional methods, since measuring quality of care via social media has other, but not less serious, limitations. Future research should explore whether social media are suitable in practice for patients, health insurers, and governments to help them judge the quality performance of professionals and organizations.

(J Med Internet Res 2014;16(2):e56) doi:10.2196/jmir.3024

KEYWORDS

social media; rating sites; patient experiences; patient satisfaction; quality of health care

Introduction

Several studies have shown significant variation in the quality of care delivered by health care providers and individual professionals [1,2]. Insight into quality of care—and especially information about the differences between providers—is important as it allows stakeholders, including consumers, health insurers, and governmental organizations such as health care inspectorates, to compare care providers and choose between them [3].

Patient centeredness is an important part of quality in health care that has gained more attention since the Institute of Medicine published its report on improving health care quality in 2001 [4]. Patient values, needs, and preferences should be respected and should guide clinical decisions. Therefore, it is essential to gain insight into quality of care from a patient's perspective. This can be achieved using traditional methods such as surveys, panels, or focus groups. Notwithstanding the potential of these strategies, they also have serious limitations [5]. First, there are several methodological challenges such as social desirability bias and selection bias [6-8]. This means that patients might give answers they think are socially accepted rather than being strictly honest (social desirability bias) or that patients who are questioned are not representative of the whole patient population (selection bias). Second, there is a time lag between the experience and the information given to the organization, insurer, patients, or health care inspectorate. Since focus groups and surveys do not allow patients to share their feedback directly after the experience, bias may occur. Third, it is difficult to reach large groups of people [9,10], and some specific groups such as ethnic minorities and people with low literacy are often not included.

Information on social media could be of value to overcome these limitations, since these new media are easy to use and are used by the majority of the population. However, people using social media are not necessarily representative of the whole population, since, for example, elderly and ethnic minorities are underrepresented in Internet use [11]. Social media are a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of user-generated content [12]. The popularity of social media can be explained by four major characteristics: they connect, create, consume, and control (online reputation) [13]. A huge and still increasing number of people use social media. For example, more than 1 billion people worldwide use Facebook, 200 million people use Twitter [14], and the number of ratings on health care rating sites has increased consistently in the past few years [15,16].

Rating sites are not a new phenomenon in our society. Many people have been using these sites to rate and find services for several years. Examples are Yelp (restaurants) and TripAdvisor (travel). In health care, rating sites allow people to share their opinion about health care providers or professionals. They are a modern way to identify what patients think and feel about health care [17]. This collection of patient experiences within health care on the Internet has been described as "crowd validation of patient experience" by Cambria et al and as a "cloud of patient experience" by Greaves et al [18,19]. Since ratings of large and complex health care services such as hospitals are hard to interpret, websites that rate individual doctors, physician-rating sites (PRSs), are a promising type of rating site. Despite resistance from the medical profession, PRSs are growing consistently [16,19].

Since many people use social media to share their experiences with health care, social media could help create transparency in the quality of health care from the patient's perspective. For example, Timian et al investigated the number of "likes" on the Facebook pages of 40 American hospitals [20]. They found that this number was negatively associated with 30-day mortality and positively with patient recommendations, which indicates a correlation between information on social media and quality of care. This example shows that social media can provide useful information about quality of care. Social networks such as Facebook and Google+, might provide information like comments, "likes", or "+1"s on the page of a hospital. Patients' experiences in health care might be shared on discussion forums or patient networks. Even microblogs, like Twitter, could function as a source of information about quality of care, although these short, unstructured messages contain minimal information [19].

The number of studies showing the information value of social media for quality of health care is growing rapidly. This has created a need for a systematic synthesis concerning the relation between social media and quality of care, its usefulness, and potential effects. Therefore, we performed a scoping review with the following goals: (1) to map the literature on the association between social media and quality of care, (2) to identify different mechanisms of this relationship, and (3) to determine a more detailed agenda for this relatively new research area.

Methods

Framework

For this study, we used the framework of Arksey and O'Malley for scoping reviews, further developed by Levac et al and Daudt et al [21-23]. A scoping study is a method to quickly map the evidence of a particular field [21]. More specifically, Mays et al defined it as follows: “to map rapidly the key concepts underpinning a research area and the main sources and types of evidence available, and can be undertaken as standalone projects in their own right, especially where an area is complex or has not been reviewed comprehensively before” [22]. It is therefore the preferred method in this study since it concerns a relatively broad issue that has not yet been clearly defined in the literature. We followed the six steps of the framework: (1) identifying the research question, (2) identifying relevant studies, (3) selecting studies, (4) charting the data, (5) collating, summarizing, and reporting the results, and (6) stakeholder consultation. The sixth step was followed throughout the study, as suggested by Daudt et al [23]. Since the research area studied in this review concerned only observational studies and is relatively new, a formal quality assessment of the included studies was not performed. However, we identified the different study designs and reported them as part of the results. The six steps of the framework will be discussed below.

Step 1: Identifying the Research Question

This scoping review focused on the association between information from patients on social media and quality of care. The research question was “What is the association between information from patients, clients, and their relatives on various types of social media and the quality of health care?” We defined information from patients, clients, and relatives on social media as any information about health care providers, health care professionals, or about the health care system in general, shared via online social media such as rating sites (rating of health care providers or professionals), (micro)blogs, social network sites, and forums. The working definition formulated by the Institute of Medicine was used to define quality of care [4]. They propose a broad definition in which good quality health care is determined by six aims: health care should be safe, effective, patient-centered, timely, efficient, and equitable.

Step 2: Identifying Relevant Studies

To identify relevant studies, we used a two-step search strategy. First, we conducted a preliminary search in PubMed to identify key articles. This step was important since this research topic is new, and little was known about relevant keywords and MeSH (Medical Subject Headings) terms. The search strategy was developed by 2 authors with expertise in performing systematic reviews (LV and RK) and further improved by an author with expertise in social media for health care (TB). The search resulted in 17 key articles.

The second step consisted of reshaping the search strategy. It was peer-reviewed by an information specialist employed at the medical library of our university hospital. A standardized list of criteria for assessing searches in the academic literature was used [24]. The final search strategy was built on four themes: social media, patient experience, quality, and health care. For every theme, thesaurus terms and text words in title and abstract were used. The themes were combined as follows: “social media” AND (“patient experience” OR “quality”) AND “health care”. Multimedia Appendix 1 shows the search strategies for the final search.

We searched four electronic databases for relevant articles: PubMed, Embase, CINAHL, and Web of Science. Since we aimed to give a broad overview of existing literature, we did not restrict the number of articles by setting limits for date of publication, type of article, or language. Additionally, we screened reference lists of included articles for relevant studies and invited several experts working in this field to share relevant articles.

Step 3: Selecting Studies

Articles were independently reviewed and scored by 2 authors (RK, LV) using title and abstract. Disagreements were discussed until consensus was reached. Finally, full texts were reviewed to determine if the articles were eligible for inclusion in the review.

For inclusion, articles should concern information from patients, clients, or their relatives on social media and the relation to quality of health care. Articles were excluded when no abstract or full text was available. A few examples of excluded articles were articles about quality improvement using a social media application (not about relation to quality of health care), articles addressing Web-based surveys about quality of health care (not about social media), and articles concerning the use of social media by medical professionals (not about information from patients, clients, or their relatives).

Step 4: Charting the Data

A data extraction form was developed by the different authors together, to ensure the approach was consistent with the research question and purpose of the scoping review. Key elements that were extracted from the articles were journal, type of study, country, type of social media application, objective(s), conclusions, and a subjective assessment of the attitude of the authors towards the relation between social media and quality of health care (“positive”, “positive with reservations”, or “negative”). To ensure that all relevant data were extracted according to the research question, all articles were assessed, and data were extracted independently by 2 researchers (LV, TB).

Step 5: Collating, Summarizing, and Reporting the Results

As proposed by Levac et al [21], we identified three distinct steps in this phase. First, we analyzed the data from the included articles and reported general characteristics. Second, the results related to the research question were described. Thus, information about the association between information on social media and quality of health care was summarized. These results were categorized by type of social media application described in the different articles. Third, the results were discussed and implications for further research, practice, and policy were described.

Step 6: Stakeholder Consultation

Professionals from the Health Care Inspectorate (the Netherlands), the Care Quality Commission (England), and several Dutch inspectorates outside of the health care sector were consulted during the process. Examples of inspectorates in other sectors were the Dutch Tax Administration

(Belastingdienst) and the Dutch Inspectorate of Education (Onderwijsinspectie). Preliminary results from this review were shared, and suggestions and advice were used to improve this scoping review.

Results

General Information

Our preliminary PubMed search resulted in 610 hits. Of these, 17 articles were labeled as key articles [3,15,19,20,25-37]. Our final search in PubMed, EMBASE, CINAHL, and Web of Science resulted in 392, 488, 55, and 73 articles respectively, totaling 1008 studies. After removing duplicates, 770 studies remained. After screening on title and abstract and reading full texts, 26 articles were included for this review. Another 3 articles were added after screening reference lists of included articles and inviting experts in the field to share relevant articles [3,15,19,20,25-49]. Figure 1 gives an overview of the study selection process.

A description of the included papers is provided in Table 1. The studies included in this review were mainly performed in the United States (n=10) or the United Kingdom (n=7). The other studies were conducted in Germany, the Netherlands, Taiwan, and Peru. Some studies did not explicitly state in which country/countries they were performed. Most articles focused on health care rating sites (n=21). Three studies concentrated on Facebook in particular, where other studies addressed social media in general. Of the 29 articles, 15 described original research. The others articles were six opinion papers, two reviews, two editorials, a news item, an essay, and two pieces of correspondence. Regarding the attitude of the authors towards the relation between social media and quality of health care, 7 authors were positive, 20 were positive but with reservations, and 2 were negative. In general, the articles could be divided according to social media application, which resulted in three groups: rating sites, Facebook, and social media in general.

Figure 1. Study selection process.

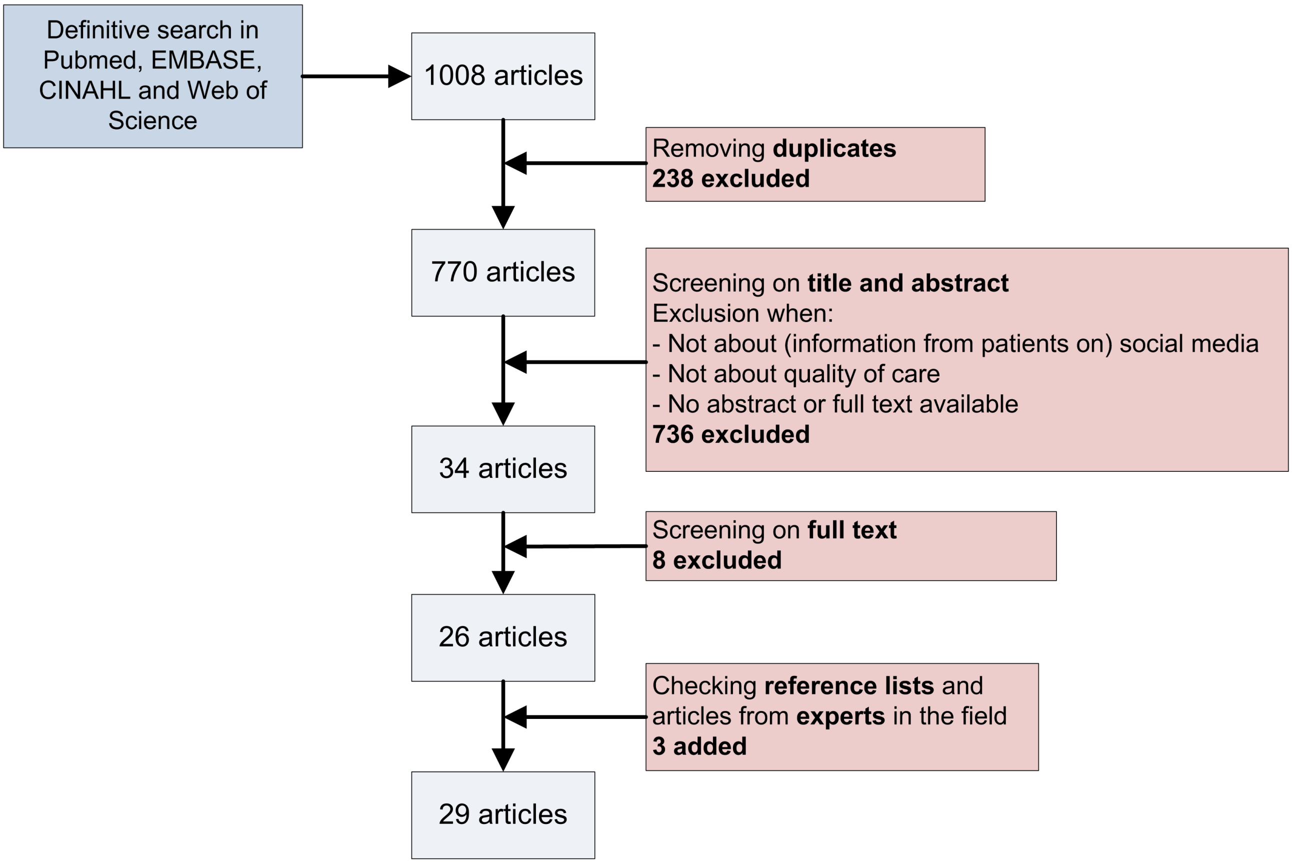


Table 1. Description of the included studies.

No.	Reference	Journal	Type of study	Country	Type of social media[a]	Attitude[b]
1	Abdul, 2011 [38]	Lancet	Correspondence	Taiwan	Facebook	+
2	Adams, 2010 [25]	Int J Med Inf	Review	Netherlands	Social media in general	+/–
3	Adams, 2011 [26]	Soc Sci Med	Original	United States, United Kingdom, Netherlands	Rating sites	+/–
4	Bacon, 2009 [39]	BMJ	Opinion paper	United Kingdom	Rating sites	+
5	Bardach, 2012 [40]	BMJ Qual Saf	Original	United States	Rating sites	+
6	Black, 2009 [41]	Inform Prim Care	Original	United States	Rating sites	+/–
7	Denecke, 2013 [42]	Methods Inf Med	Editorial	n/a	Social media in general	+/–
8	Emmert, 2012 [27]	Methods Inf Med	Original	Germany	Rating sites	+/–
9	Emmert, 2013 [43]	J Med Internet Res	Review	n/a	Rating sites	+/–
10	Galizzi, 2012 [44]	BMJ Open	Original	United Kingdom	Rating sites	+/–
11	Gao, 2012 [15]	J Med Internet Res	Original	United States	Rating sites	+/–
12	Greaves, 2012[c][28]	BMJ Qual Saf	Original	United Kingdom	Rating sites	+
13	Greaves, 2012 [29]	Arch Intern Med	Original	United Kingdom	Rating sites	+/–
14	Greaves, 2012 [30]	J Med Internet Res	Original	United Kingdom	Rating sites	+/–
15	Greaves, 2013 [19]	BMJ Qual Saf	Opinion paper	United Kingdom	Social media in general	+/–
16	Hammond, 2008 [45]	Guidelines in Practice	Opinion paper	n/a	Rating sites	+
17	Kadry, 2011 [31]	J Med Internet Res	Original	United States	Rating sites	+/–
18	Lagu, 2010 [32]	J Gen Intern Med	Original	United States	Rating sites	+/–
19	Lopez, 2012 [46]	J Gen Intern Med	Original	United States	Rating sites	+/–
20	McCartney, 2009 [47]	BMJ	Opinion paper	United Kingdom	Rating sites	–
21	Reimann, 2010 [33]	BMC Health Serv Res	Original	n/a	Rating sites	+/–
22	Rozenblum, 2013 [48]	BMJ Qual Saf	Editorial	n/a	Social media in general	+
23	Segal, 2012 [34]	J Med Internet Res	Original	United States	Rating sites	+/–
24	Strech, 2011 [35]	J Med Internet Res	Opinion paper	n/a	Rating sites	+/–
25	Tanne, 2013 [36]	BMJ	News item	United States	Rating sites	+/–
26	Tello, 2013 [49]	Am J Med Qual	Correspondence	Peru	Facebook	–
27	Thielst, 2011 [37]	Front Health Serv Manage	Opinion paper	United States	Social media in general	+
28	Timian, 2013 [20]	Am J Med Qual	Original	United States	Facebook	+
29	Trigg, 2011 [3]	J Health Serv Res Policy	Essay	n/a	Rating sites	+/–

[a] Rating sites: various types of health care rating sites like physician rating sites or hospital rating sites.

[b] A subjective assessment of the authors' attitude towards the relation between social media and quality of health care (+: positive, +/–: positive with reservations, –: negative).

[c] We refer to three different papers by Greaves et al in 2012: see citations [28-30].

Association Between Types of Social Media and Quality of Care

Rating Sites

Association Between Ratings and Quality of Care

Most identified studies (21/29) concerned the association between ratings on rating sites and quality of care. Table 2 shows the correlations that have or have not been shown in various original studies. Greaves et al demonstrated a correlation between Web-based patient ratings of hospitals (on NHS Choices) and conventional surveys of patient experiences. Furthermore they showed a relationship between these ratings and objective measures of quality, including readmission rates, mortality, and infection rates. They investigated the same associations with ratings of family physician practices on NHS choices. These ratings are moderately associated with measures of patient experience and weakly with clinical quality [28-30]. A study into the rating of physicians on the RateMDs website suggests that these ratings correlate positively with physician quality, measured by board certification, education, and malpractice claims. The ratings were based on a small number of reviews, and most rating variation reflected evaluations of punctuality and staff. The authors concluded that further research is needed into the correlation between ratings and clinical outcomes [15]. Segal et al found that while the total number of reviews correlated with surgeon volume (as a proxy for surgeon quality), the actual rating value did not [34]. Bardach et al describe a correlation between hospital scores on Yelp (a commercial rating website) and a more traditional measure of patient experience, the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) scores. Furthermore, they found correlations between Yelp scores and mortality and readmission rates. According to the authors, their data suggests that consumers posting ratings on Yelp may observe aspects of care related to important patient outcomes [40].

Next to these original studies, three other studies argue that information from rating sites reflects quality of care. Hammond (2008), broadcaster and general practitioner, thinks the safety is in numbers: “If you can’t find 200 patients who approve of what you’re doing, you’re in the wrong job!” [45]. Bacon, physician and shareholder of the rating site, iWantGreatCare, argues that rating sites provide valid, detailed, and timely feedback that is needed to efficiently measure quality and satisfaction. Therefore, professionally responsible rating sites will increase standards of care but only for those organizations and doctors that think the experience of the patient is as important as excellent clinical outcome [39]. Finally, Trigg, researcher in the field of health care and social care, focuses on the use of PRSs by patients. She states that the increasing use of these sites suggests that patients seek new ways to give feedback on care providers. Understanding the reasons for use of PRSs can give insight into how the information can be used for quality improvement [3].

In contrast with the studies mentioned above, various studies suggest that patient ratings are not, or not yet, useful enough to give insight into the quality of care for the following reasons.

Table 2. Correlations between information from social media and measures of quality[a].

		Measure of quality								
Article	Info from social media	Patient experiences	Readmission rates (different measures used)	Mortality (different measures used)	Board certification	Education	Malpractice claims	Infection rates	Clinical quality indicators	Surgeon volume
Bardach, 2012 [40]	Hospital rating	+	+	+/–						
Gao, 2012 [15]	Physician rating				+	+	+/–			
Greaves, 2012 [28,29]	Hospital rating	+	+	+/–				+		
Greaves, 2012 [30]	Family physician rating	+							+/–	
Segal, 2012 [34]	No. of reviews									+
	Rating value									–
Timian, 2013 [20]	No. of “likes”	+		+						

[a] This table presents the correlations/associations as stated by the authors in the various papers (+: there is correlation, +/–: correlation is weak or not found for all aspects, –: there is no correlation).

Partial Quality Measurement

Reimann et al performed a study to investigate the extent to which English- and German-language PRSs represented different dimensions of patient experience and satisfaction, determined by a systematic review. They identified 13 dimensions in three categories: characteristics of encounter between doctor and patient (eg, trust), organizational aspects (eg, accessibility), and overarching assessment categories (eg, general satisfaction). They found that none of the 21 investigated PRSs represented all 13 dimensions. However, the three most visited German sites represented between 8 and 11 dimensions. The three most trafficked English-language PRSs represented between 5 and 6 dimensions. Specifically the dimensions communication skills and information/advice were missing [33].

Influence of Patient Characteristics on Ratings

Emmert et al state that patient satisfaction and outcome measures on PRSs are not risk-adjusted [27], although research has shown that patient satisfaction results are influenced by age, education, and health status [6]. Also, Galizzi et al found that subjects who give feedback on doctor-ranking websites are unlikely to be representative of the overall patient pool. This indicates that it is important to look at user characteristics when interpreting results from doctor-rating sites [44].

Positive Sentiment of Ratings

Four original studies analyzed the content of reviews on rating sites. Black et al analyzed 16,703 reviews on 6101 providers in the United States. They found that online ratings were largely positive [41]. This was also found by Lagu et al for physician ratings. The study also identified narratives that appeared to be written by physicians themselves [32]. Kadry et al performed an analysis of 4999 online physician ratings. They concluded that most patients give physicians a favorable rating [31]. Finally, Lopez et al analyzed 712 reviews of primary care physicians. The majority of these reviews was positive [46].

Factors Other Than Quality of Care on Ratings

Adams et al performed an analysis of four share-your-experience websites in three countries, supplemented by interviews with stakeholders from the Netherlands (website developers, hospitals, insurers, and members of the Dutch Health Inspectorate). Their results show that the sharing of experiences by patients is not automatic but encouraged by website creators who have their own purpose with these posts [26]. Lopez et al conclude that “patient reviews are affected by more aspects of care than the patient-physician interaction only. Accessibility, convenience and staff also play a role” [46].

Low Number of Reviews

A study looking at the ratings of 500 randomly selected urologists on 10 rating websites showed an average of 2.4 ratings per doctor. According to Tanne et al, this indicates that these sites need more reviews to make them more reliable [36].

Ratings and the Potential Harm to a Physician’s Reputation

Strech et al addressed the ethical discussion around the basic concept of PRSs. They conclude that the potential harms for physicians that can result from PRSs (financial and psychological) need to be contained without limiting the potential benefits for patients with respect to health, health literacy, and equity [35]. McCartney, a general practitioner, goes one step further. She argues against the use of rating sites in health care because “it is a non-evidence-based intervention with potentially damaging strings attached”. For example, some medical work, like child-protection and psychiatry, has the constant potential for conflict. Also, factors like socioeconomic status might influence satisfaction with general practice services [47]. Emmert et al performed a systematic review of the literature about physician rating sites. They conclude that rating sites are gaining more attention in research and mention several shortcomings of these sites from literature. Examples include the fact that it is often not possible to relate anonymous feedback to specific incidents, making it unlikely that care providers can learn from the comments. Also, anonymous ratings makes it easy to abuse these sites, which might lead to defamation of professionals or misinformation to patients [43].

Facebook

Three articles focused in particular on the social network site Facebook. A study by Timian et al, involving 40 hospitals in the United States, found that the number of “Likes” on the Facebook page of the hospitals had a negative association with the 30-day mortality rate and a positive association with patient recommendation [20]. Tello et al commented on this article by stating that measuring quality of care with Facebook likes in Peru is confronted with several barriers. For example, people on Facebook are hardly representative of the patient population since only a small proportion of Peruvians, and mostly younger people, use Facebook [49]. Next to this, Abdul describes how Facebook enabled collaboration between stakeholders in emergency-medicine policy in Taiwan. An active discussion on a Facebook group about overcrowded emergency rooms was followed by the Minister of Health’s involvement, which eventually led to health care reforms in the country [38].

Social Media in General

Five articles focused on social media in general, without handling one type in particular. Adams et al performed a literature and Web review on the reliability of online health information in light of the increasingly popular Web 2.0. They state that issues about reliability, like disclosure of authorship and privacy, should not easily be dismissed. Therefore, caution is required when newly popular Web applications are used for health purposes [25]. Thielst discusses the use of social media in health care. She argues that social media platforms are a cheap way for health care organizations to hear the voice of patients and get feedback on their care [37]. Furthermore, an editorial by Denecke et al reports that information from medical social media could provide a new source of information. For example, patient stories on discussion forums could enable earlier detection of adverse drug effects [42]. Rozenblum et al also emphasize the growing importance of patients’ experience acquired from social media. They think that this information will complement traditional patient surveys and will help identify poor care and outstanding care [48]. Finally, Greaves et al describe the possibility of using the “cloud of patient experience” on the Internet for detection of poor quality care. They provide advantages and disadvantages of different sources of information (eg, rating sites, patient forums, social networks) and name several technical and logistic limitations for using and processing information about quality of care from social media. The authors suggest the comparison between conventional measures of patient experience with information from the online cloud of patient experiences (after collection and processing) in future research [19].

Discussion

Principal Findings

This review showed that, although literature about the topic is limited, several studies indicate a relation between information on social media and quality of health care. Interestingly, most of these studies concern rating sites. An association was found for ratings of whole organizations as well as for individual physicians, although different measures for quality of care were used. These findings show that social media, and especially rating sites, could be a fast and efficient way to gather information about quality of care. However, several disadvantages of using social media also exist. In this discussion, we will put our findings in perspective.

This scoping review identified studies in which subjective ratings were not only correlated to subjective measures of quality but also to objective measures. Patient rating of a certain hospital is likely to correlate with patient recommendation. Since patients are not likely to have insight into hospital mortality or infection rates, the associations shown by for example Timian et al [20] and Greaves et al [29] are remarkable. The fact that these patient ratings correlate to (aspects of) quality of health care might make expensive, traditional measures of patient experiences unnecessary in the future. However, associations in the included studies were shown only at one point in time. Research is also needed into the predictive value of ratings over time [19]. This feature is especially of interest for supervisory bodies such as health care inspectorates. Therefore, it is important to perform studies with a longitudinal design. When looking at different purposes for gathering information about quality of health care, social media might be useful as a predictor for low quality health care, which has already been shown by Google flu trends. This sophisticated tool from Google analyzes health care-related search queries from people worldwide in their search engine. Because there is a close relationship between the number of people searching for influenza-related topics and those who have influenza symptoms, this tool can predict flu outbreaks much faster than conventional surveillance [50]. Furthermore, new techniques are being developed to analyze unstructured data about the quality of health care on the Internet. Greaves et al showed that sentiment analysis of patients' comments about their health care is possible and reasonably accurate [51].

In contrast with these findings, the articles included in this study also identified several drawbacks concerning rating sites. These include the fact that rating is often anonymous and as a result, rating values are not risk-adjusted and are vulnerable to fraud. A health care professional can, for example, rate him/herself or colleagues. Also, ratings are often based on only a few reviews and are predominantly positive. Furthermore, people providing feedback on health care via social media are presumably not always representative of the patient population. Further, reviews from patients are influenced by other factors than quality of care. Not only the patient-physician relation is rated but issues like accessibility play a role too. Therefore, several authors suggest that information from social media should be used with caution. Also, several examples exist in the Netherlands where information on new media is biased. A popular Dutch opinion website has influenced several polls by encouraging people to vote for a certain answer [52]. We are also aware of a case where media attention around a poorly performing hospital elicited more positive reviews shared by people who wanted to stand up for this hospital. This shows that groups of people can purposely influence information on social media. Another important issue is that there are many organizational differences between rating sites that can influence the content presented. Examples are the presence or absence of editors that check ratings before they appear on the site and the possibility for doctors to share their views on comments [53]. Also, it should be realized that rating sites can be owned by stakeholders that may have conflicts or interests. Sometimes these sites are organized by the government, such as NHS Choices, and sometimes these sites can be privately owned or owned by patient federations such as ZorgkaartNederland in the Netherlands.

Limitations

Our study has some limitations. It is possible that publication bias has affected our results since studies without significant results are often not published. However, we found many articles that also discussed the negative aspects of using social media to gain insight into quality of health care. This suggests that influence of publication bias was minimal. Furthermore, our study was restricted to literature from online databases. Future studies might consider inclusion of grey literature.

Conclusion

Social media and rating sites in particular are an interesting new source of information about quality of care. However, this new source should, at least for now, be used to complement traditional methods, since measuring quality of care via social media has other, but not less serious, limitations [28]. Future research should focus on comparing objective traditional measures of quality with subjective information from social media, which has also been suggested by other authors [17,54]. This will provide more evidence on the association between the two approaches. Furthermore, this scoping review provides a basis for a more systematic review of the literature, which can give a more definite answer about how information from social media can be used to assess quality of health care.

Acknowledgments

This study was performed in the context of a project funded by ZonMw, the Netherlands Organisation for Health Research and Development. The publication fee was reimbursed by the Netherlands Organisation for Scientific Research (NWO).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategies for the different online databases.

[PDF File (Adobe PDF File), 70KB - jmir_v16i2e56_app1.pdf]

References

1. Wennberg J, Gittelsohn A. Small area variations in health care delivery. Science 1973 Dec 14;182(4117):1102-1108. [Medline: 4750608]
2. Wennberg JE. Practice variations and health care reform: connecting the dots. Health Aff (Millwood) 2004;Suppl Variation:VAR140-VAR144 [FREE Full text] [doi: 10.1377/hlthaff.var.140] [Medline: 15471778]
3. Trigg L. Patients' opinions of health care providers for supporting choice and quality improvement. J Health Serv Res Policy 2011 Apr;16(2):102-107. [doi: 10.1258/jhsrp.2010.010010] [Medline: 20870690]
4. Berwick DM. A user's manual for the IOM's 'Quality Chasm' report. Health Aff (Millwood) 2002;21(3):80-90 [FREE Full text] [Medline: 12026006]
5. Willis CD, Evans SM, Stoelwinder JU, Cameron PA. Measuring quality. Aust Health Rev 2007 May;31(2):276-281. [Medline: 17470049]
6. Hekkert KD, Cihangir S, Kleefstra SM, van den Berg B, Kool RB. Patient satisfaction revisited: a multilevel approach. Soc Sci Med 2009 Jul;69(1):68-75. [doi: 10.1016/j.socscimed.2009.04.016] [Medline: 19446942]
7. Giebel GD, Groeben N. Social desirability in the measuring of patient satisfaction after treatment of coloproctologic disorders: on shortcomings of general bipolar satisfaction scales for quality management. Langenbecks Arch Surg 2008 Jul;393(4):513-520. [doi: 10.1007/s00423-008-0310-x] [Medline: 18330596]
8. Mazor KM, Clauser BE, Field T, Yood RA, Gurwitz JH. A Demonstration of the Impact of Response Bias on the Results of Patient Satisfaction Surveys. Health Serv Res 2002 Oct;37(5):1403-1417. [doi: 10.1111/1475-6773.11194]
9. Levine D, Madsen A, Wright E, Barar RE, Santelli J, Bull S. Formative research on MySpace: online methods to engage hard-to-reach populations. J Health Commun 2011 Apr;16(4):448-454. [doi: 10.1080/10810730.2010.546486] [Medline: 21391040]
10. Lovaglio PG. Benchmarking strategies for measuring the quality of healthcare: problems and prospects. ScientificWorldJournal 2012;2012:606154 [FREE Full text] [doi: 10.1100/2012/606154] [Medline: 22666140]
11. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
12. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
13. Hoffman DL, Novak TP. Why Do People Use Social Media? Empirical Findings and a New Theoretical Framework for Social Media Goal Pursuit. SSRN Journal 2012 Jan 17:1-33 [FREE Full text] [doi: 10.2139/ssrn.1989586]
14. Marketingfacts Stats Dashboard: social media marketing. URL: http://www.marketingfacts.nl/statistieken/social-media-marketing/ [accessed 2013-09-03] [WebCite Cache ID 6JM99OZ8g]
15. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
16. Greaves F, Millett C. Consistently increasing numbers of online ratings of healthcare in England. J Med Internet Res 2012;14(3):e94 [FREE Full text] [doi: 10.2196/jmir.2157] [Medline: 22742977]
17. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
18. Cambria E, Hussain A, Havasi C, Eckl C, Munro J. Towards crowd validation of the UK national health service. WebSci10 2010:1-5.
19. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]
20. Timian A, Rupcic S, Kachnowski S, Luisi P. Do patients "like" good care? measuring hospital quality via Facebook. Am J Med Qual 2013;28(5):374-382. [doi: 10.1177/1062860612474839] [Medline: 23378059]
21. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
22. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
23. Daudt HM, van Mossel C, Scott SJ. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol 2013;13:48 [FREE Full text] [doi: 10.1186/1471-2288-13-48] [Medline: 23522333]
24. McGowan J, Sampson M, Lefebvre C. An evidence based checklist for the peer review of electronic search strategies (PRESS EBC). Evidence Based Library and Information Practice 2010;5(1):149-154.
25. Adams SA. Revisiting the online health information reliability debate in the wake of "web 2.0": an inter-disciplinary literature and website review. Int J Med Inform 2010 Jun;79(6):391-400. [doi: 10.1016/j.ijmedinf.2010.01.006] [Medline: 20188623]

26. Adams SA. Sourcing the crowd for health services improvement: The reflexive patient and "share-your-experience" websites. Soc Sci Med 2011 Apr;72(7):1069-1076. [doi: 10.1016/j.socscimed.2011.02.001] [Medline: 21414701]
27. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. Methods Inf Med 2012;51(2):112-120. [doi: 10.3414/ME11-01-0045] [Medline: 22101427]
28. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012 Jul;21(7):600-605. [doi: 10.1136/bmjqs-2012-000906] [Medline: 22523318]
29. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Web-based patient ratings and objective measures of hospital quality. Arch Intern Med 2012 Mar 12;172(5):435-436. [doi: 10.1001/archinternmed.2011.1675] [Medline: 22331980]
30. Greaves F, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, et al. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. J Med Internet Res 2012;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]
31. 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]
32. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
33. 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]
34. Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? J Med Internet Res 2012;14(2):e50 [FREE Full text] [doi: 10.2196/jmir.2005] [Medline: 22491423]
35. Strech D. Ethical principles for physician rating sites. J Med Internet Res 2011;13(4):e113 [FREE Full text] [doi: 10.2196/jmir.1899] [Medline: 22146737]
36. Tanne JH. Doctor-rating websites base their reports on only a few patient reviews. BMJ 2013;346:f295. [Medline: 23321414]
37. Thielst CB. Social media: ubiquitous community and patient engagement. Front Health Serv Manage 2011;28(2):3-14. [Medline: 22256506]
38. Abdul SS, Lin CW, Scholl J, Fernandez-Luque L, Jian WS, Hsu MH, et al. Facebook use leads to health-care reform in Taiwan. Lancet 2011 Jun 18;377(9783):2083-2084. [doi: 10.1016/S0140-6736(11)60919-7] [Medline: 21684378]
39. Bacon N. Will doctor rating sites improve standards of care? Yes. BMJ 2009;338:b1030. [Medline: 19293223]
40. Bardach NS, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar;22(3):194-202. [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]
41. 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]
42. Denecke K, Brooks E. Web science in medicine and healthcare. Methods Inf Med 2013;52(2):148-151. [Medline: 23508344]
43. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
44. Galizzi MM, Miraldo M, Stavropoulou C, Desai M, Jayatunga W, Joshi M, et al. Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London. BMJ Open 2012;2(6):e001493 [FREE Full text] [doi: 10.1136/bmjopen-2012-001493] [Medline: 23148340]
45. Hammond P. Out of hours. Online rating will allow for increased patient feedback. Guidelines in Practice 2008;11(5):72-72.
46. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
47. McCartney M. Will doctor rating sites improve the quality of care? No. BMJ 2009;338:b1033. [Medline: 19293224]
48. Rozenblum R, Bates DW. Patient-centred healthcare, social media and the internet: the perfect storm? BMJ Qual Saf 2013 Mar;22(3):183-186. [doi: 10.1136/bmjqs-2012-001744] [Medline: 23378660]
49. Tello S, Torres MF, Monroe FJ, Segura ER. Application of Facebook "likes" as an indicator of quality in health care in public hospitals: barriers and opportunities in Peru. Am J Med Qual 2013;28(4):358. [doi: 10.1177/1062860613488435] [Medline: 23687237]
50. 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]
51. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
52. Oranje favoriet in Duitse poll na inmenging GeenStijl. 2010. URL: http://www.nu.nl/internet/2285468/oranje-favoriet-in-duitse-poll-inmenging-geenstijl.html [accessed 2013-09-03] [WebCite Cache ID 6JM8gKjbp]

53. 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]
54. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]

Abbreviations

HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
MeSH: medical subject headings
NHS: National Health Service
NWO: Netherlands Organisation for Scientific Research
PRS: physician-rating sites
ZonMw: The Netherlands Organisation for Health Research and Development

Edited by G Eysenbach; submitted 14.10.13; peer-reviewed by F Greaves; comments to author 07.01.14; revised version received 17.01.14; accepted 19.01.14; published 20.02.14.

Please cite as:
Verhoef LM, Van de Belt TH, Engelen LJLPG, Schoonhoven L, Kool RB
Social Media and Rating Sites as Tools to Understanding Quality of Care: A Scoping Review
J Med Internet Res 2014;16(2):e56
URL: http://www.jmir.org/2014/2/e56/
doi:10.2196/jmir.3024
PMID:24566844

Original Paper

Growing a Professional Network to Over 3000 Members in Less Than 4 Years: Evaluation of InspireNet, British Columbia's Virtual Nursing Health Services Research Network

Noreen Frisch[1], RN, PhD; Pat Atherton[1], BCom, MPA; Elizabeth Borycki[1,2], RN, PhD; Grace Mickelson[3], RN, MA; Jennifer Cordeiro[4], BAA; Helen Novak Lauscher[4], PhD; Agnes Black[5], RN, MPH

[1]School of Nursing, Faculty of Human and Social Development, University of Victoria, Victoria, BC, Canada

[2]School of Health Information Science, Faculty of Human and Social Development, University of Victoria, Victoria, BC, Canada

[3]Provincial Health Services Authority, Vancouver, BC, Canada

[4]eHealth Strategy Office, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[5]Providence Health Care, Vancouver Coastal Health, Vancouver, BC, Canada

Corresponding Author:
Noreen Frisch, RN, PhD
School of Nursing
Faculty of Human and Social Development
University of Victoria
PO Box 1700, STN CSC
Victoria, BC, V8W 2Y2
Canada
Phone: 1 250 721 7953
Fax: 1 250 721 6231
Email: nfrisch@uvic.ca

Abstract

Background: Use of Web 2.0 and social media technologies has become a new area of research among health professionals. Much of this work has focused on the use of technologies for health self-management and the ways technologies support communication between care providers and consumers. This paper addresses a new use of technology in providing a platform for health professionals to support professional development, increase knowledge utilization, and promote formal/informal professional communication. Specifically, we report on factors necessary to attract and sustain health professionals' use of a network designed to increase nurses' interest in and use of health services research and to support knowledge utilization activities in British Columbia, Canada.

Objective: "InspireNet", a virtual professional network for health professionals, is a living laboratory permitting documentation of when and how professionals take up Web 2.0 and social media. Ongoing evaluation documents our experiences in establishing, operating, and evaluating this network.

Methods: Overall evaluation methods included (1) tracking website use, (2) conducting two member surveys, and (3) soliciting member feedback through focus groups and interviews with those who participated in electronic communities of practice (eCoPs) and other stakeholders. These data have been used to learn about the types of support that seem relevant to network growth.

Results: Network growth exceeded all expectations. Members engaged with varying aspects of the network's virtual technologies, such as teams of professionals sharing a common interest, research teams conducting their work, and instructional webinars open to network members. Members used wikis, blogs, and discussion groups to support professional work, as well as a members' database with contact information and areas of interest. The database is accessed approximately 10 times per day. InspireNet public blog posts are accessed roughly 500 times each. At the time of writing, 21 research teams conduct their work virtually using the InspireNet platform; 10 topic-based Action Teams meet to address issues of mutual concern. Nursing and other health professionals, even those who rated themselves as computer literate, required significant mentoring and support in their efforts to adopt their practice to a virtual environment. There was a steep learning curve for professionals to learn to work in a virtual environment and to benefit from the available technologies.

Conclusions: Virtual professional networks can be positioned to make a significant contribution to ongoing professional practice and to creating environments supportive of information sharing, mentoring, and learning across geographical boundaries. Nonetheless, creation of a Web 2.0 and social media platform is not sufficient, in and of itself, to attract or sustain a vibrant community of professionals interested in improving their practice. Essential support includes instruction in the use of Web-based activities and time management, a biweekly e-Newsletter, regular communication from leaders, and an annual face-to-face conference.

(J Med Internet Res 2014;16(2):e49) doi:10.2196/jmir.3018

KEYWORDS

social networking; social media; nursing; health services; research; education

Introduction

InspireNet

In 2009, "InspireNet" (Innovative Nursing Services and Practice Informed by Research and Evaluation Network) was launched with the purpose of increasing nurses' capacity for and interest in health services research in British Columbia (BC), Canada. Funded as part of the BC Nursing Research Initiative through the Michael Smith Foundation for Health Research, the network team accepted a mission to serve nurses and other health professionals in diverse and geographically dispersed settings. Operating a network virtually, with the ability to connect people both synchronously and asynchronously, was the only viable option to reach, engage, and support nurse clinicians, managers, educators, researchers, and students, given the nature of their lives and work. InspireNet was, to the planning team's knowledge, the first health care professionals' network in BC that attempted to conduct its work almost exclusively through Web 2.0, social media, and Web conferencing technologies, and likely one of the first of its kind in Canada. For the purpose of this paper, Web 2.0 is used in a broad sense to represent interactive forms of connectivity that permit many-to-many communications in both synchronous and asynchronous timeframes.

In 2012, preliminary findings documenting the network's growth and the experience of one of our topic-based Action Teams were presented [1]. In this current paper, we present evaluative data gathered over a three-year period on network growth, the parts of the network that support frequent use of the network platform, the establishment of password-protected electronic communities of practice, or eCoPs, (some of which are "open" to all network members and others that are "closed" and available only to specific members of working teams), and a summary of data obtained from our surveys and interviews, indicating the reasons individuals joined and participate in the network.

Review of the Literature

There is a rapidly developing body of literature on the use of Web 2.0 and social media in health. A Medline search indicates the number of publications have more than doubled over the past five years with more than 3500 related publications. Many of the papers discuss how electronic tools assist individuals to support their health and health conditions. Another group of papers focuses on the use of social media to enhance communication between health care providers and patients, and among health care providers. A gap was identified as a small but emerging area of research at the intersection of social media, Web 2.0, and health professional practice; there were 93 papers published, 67 of them within the last 5 years [1]. This research area focuses on how health care professionals use electronic tools to connect with and learn from one other and seeks to understand the impact of these professional virtual connections. Researchers have reported that social media can be an effective tool to establish mentoring relationships [2], serve as a teaching tool for continuing professional education [3,4], provide a means to share professional ideas [5,6], and serve as a way for professional organizations to reach out to members/potential members to broaden their scope [7]. In 2009, writers for the Medical Library Association News commented that "...never before have there existed so many opportunities to 'meet' other professionals...Social networking enables meetings and collaborations on a level that has never existed before…" [8]. A recent study conducted in the Netherlands indicates that nearly 60% of health care providers use social media professionally and that their motives for using it include increasing their knowledge, efficiencies, and communication with both patients and colleagues [9]. InspireNet's use of Web 2.0 and social media adopted these ideas and extended the idea of professional networking through development of a formal virtual network, inviting nurses and other health care professionals to join and work collaboratively on issues or topics of personal interest. Further, the network espoused the view that eCoPs could provide a means for productive work to take place. InspireNet was developed as a network of electronic communities of practice connected to each other through a virtual platform with a common vision of supporting and using health services research.

Communities of practice, or CoPs, were initially described by authors influenced by social learning theory. Wenger identifies that learning takes place within social relationships and is not simply based on knowledge development [10]. In a CoP, people who share an interest in a topic, have a common concern, or wish to solve a problem, collaborate and share ideas and experiences, expanding their knowledge and expertise. A CoP takes time to develop, yet sustains itself as an entity working toward a common goal. There are "life-cycle" phases of CoPs describing a progression from being an informal group to an actively engaged community: (1) potential—an informal group, (2) coalescing—establishing a group identity, (3) maturing—actively working together to a shared goal, (4) stewardship—sustaining momentum, and (5) transformation—having accomplished a goal, the members then identify a new goal or disband [11]. eCoPs are communities making use of electronic, or virtual, platforms. Experiences of those using eCoPs demonstrate that factors such as voluntary involvement, distributed leadership, shared identity, transparency, accessibility, and being problem-focused contribute to their success [12]. Further, in a recent article published in this journal, the author commented that building and sustaining effective eCoPs requires an “enabler” and “strategic community management” [13].

One aspect of CoPs that has been described is the division of members into “core” and “peripheral” categories. These terms were first used by Lave and Wenger when they described those who were entering learning communities as “new learners”, standing in the background [14]. Over time, those peripheral members become more involved, becoming “core” members. As CoPs entered the Web-based world, authors noted that eCoPs seem to experience a similar dynamic. In eCoPs especially, the term “peripheral members”, or “lurkers”, emerged as referring simply to those who “look in” to the eCoP but are not active. At least one group of authors cautions against the use of this derogatory term, as “lurkers” may be not only passively benefitting from the eCoP activities, they may also become quite active or even core members when the work on the topic of interest speaks directly to them [15]. In Web-based eCoPs, peripheral members may be a majority of members, yet their motivations for membership, their commitment to the eCoP, and the benefits they accrue are difficult to measure.

Background on InspireNet

Overview

InspireNet was designed to make full use of Web technologies with the goals to (1) support the professional development of nurses and other professionals in the area of health services research, (2) support the development of research teams, and (3) enhance knowledge translation within health care systems. There is no monetary cost for membership in the network and membership is completely voluntary. InspireNet accepted the notion from the beginning that there would be core and peripheral members and also that members might experience periods of greater or lesser participation, depending on multiple factors such as time constraints and the topic of discussion. InspireNet’s virtual platform is comprised of several components: a Web 2.0 website, Web conferencing, and social media. Each is described below.

The Website

The InspireNet website [16] was designed to be interactive and engaging, providing a platform for members to contribute to website content without advanced skills or specialized training. Figure 1 provides a screen capture of the site’s homepage. Figure 2 illustrates the website map. See Multimedia Appendix 1 for website details.

Figure 1. InspireNet home page.

Welcome to InspireNet.ca!

Welcome message

We are a network of 3,000 researchers, practitioners, policy-makers, educators and students working together to improve nursing health services in British Columbia, Canada. To explore this website, use the navigation links on the left hand side and the Search tool. Join InspireNet to keep up to date on fresh web content and to join the discussion and Action Teams.

Connecting People, Ideas and Solutions.

InspireNet Annual Evaluation Survey

InspireNet members - we need your help!

Please take a few minutes to complete our annual survey. (You must be logged in to the website to access the survey).

Thank you in advance for your help.

►**Survey:** http://www.inspirenet.ca/fall2013-survey

Figure 2. InspireNet site map.

Web Conferencing

Supported by Cisco WebEx, Web conferencing provides InspireNet members the ability to connect in real time via their own computers, eliminating the need for travel and its associated burden. Action Teams use WebEx to facilitate for their team members free learning activities (webinars), discussions exploring potential research projects, and project development meetings. Closed Teams use WebEx in project development meetings, in conjunction with the website, for document-sharing and asynchronous discussion via forums within their eCoP. Web conferencing provides the ability for all teams to access guest speakers and consultants for lectures and discussions with team members. InspireNet teams have interacted with guests from across Canada and from Australia, Europe, and the United States. Action Teams' webinars are recorded and archived for viewing by team members. At the time of writing, 79 webinars had been held with approximately 4000 views; InspireNet's experience indicates that most views occur asynchronously via recordings, at members' leisure.

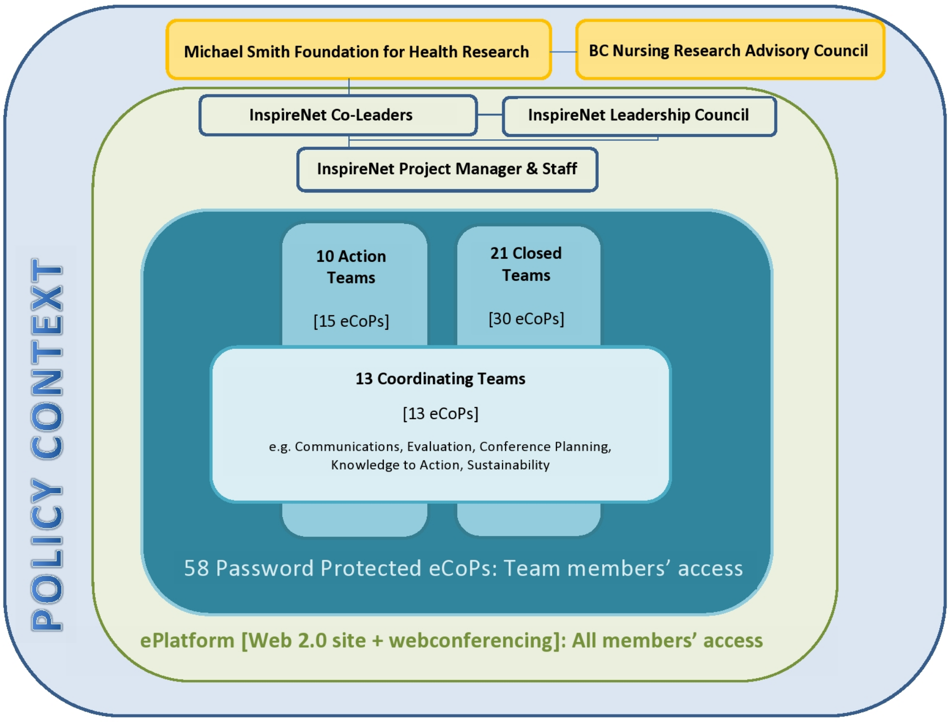
Social Media

InspireNet leverages the power of social media to link relevant content with members, to grow membership, and to raise awareness about nursing health services research through an active Twitter feed (@InspireNetBC), a Facebook page (InspireNet), and a YouTube channel (InspireNetBC). Members are encouraged to include their LinkedIn account as a link in their database profile to advance connections using that platform.

Distributed Leadership as a Working Model

InspireNet does not rely on technology alone to support its members. A distributed leadership model was developed to promote and sustain member collaboration (see Figure 3). The distributed leadership model is innovative and has been a key part of the success of InspireNet. This model describes that having professionals work together creates collaborative advantages helpful to all [17]. The management team consists of three individuals—two co-leaders, one from clinical practice and one from academia, and a professional network manager. The network manager is retained at 0.8 full-time equivalent (FTE); the network co-leaders provide in-kind support to the network through their respective organizations that have agreed to serve as the network's host institutions. With the exception of technical and scheduling support work carried out by the manager and a 0.2 FTE administrative support, all contributions to the network are voluntary. Coordinating teams' members are professionals who volunteer to support the infrastructure and the work of the network. As is the nature of content on the World Wide Web generally and consistent with the theory on CoPs, InspireNet teams regulate themselves and contribute out of a desire to make a difference in their work and to support a common goal. Figure 3 provides an illustration of the relationship between the teams, where the Coordinating Teams support the work of Action and Closed Teams and the general membership.

Figure 3. Network framework.



Methods

Evaluation Framework

The network evaluation framework includes methods of retrieving data to document network activities, to understand the benefits and challenges of establishing a virtual professional network, and to learn what factors are valued and needed to support network growth. The range of evaluative methods for documenting and tracking members' perceptions and usage of InspireNet are described in detail below.

Network Growth

Monthly metrics are tabulated and reported as a dashboard and uploaded to the website for public access, demonstrating network growth generally, as well as in Action and Closed Teams, and milestone achievements.

Network Use

Google Analytics is used by the network manager for the dashboard, reporting on webpage hits; periodically, more in-depth analysis is done to examine trends in website traffic in order to inform future website content decisions and to learn more about members' use of the website. Additionally, Drupal provides metrics on the number of reads on each webpage for "on-the-fly" updates; these metrics are publicly available to any website visitor.

Member Surveys

Network members have been surveyed twice. Online surveys addressed members' satisfaction measured against achievements toward the network's goals. Surveys were administered after 17 months of operation (Spring 2011) and after 30 months of operation (Spring 2012). Data were analyzed through descriptive statistics.

Interviews

Interviews were conducted with Action Team leaders and members between 22 and 24 months of network operation (September to December 2011). Participants were selected via purposive sampling from those who indicated interest in participating in an interview. A protocol was developed to question interviewees about their involvement in InspireNet/Action Teams and their perceptions of the network's activities and successes in reaching its goals. Interviews were 30-60 minutes, conducted by a graduate student via computer using Audacity (open source audio recording software), and transcribed verbatim. A coding scheme based on the framework of the InspireNet goals was developed by the evaluation team and NVivo (qualitative data analysis software) was used to code the 23 transcripts. Coding was done by four evaluation team members and a reliability check was conducted to ensure consistency across the transcripts. Once coded, the content was analyzed and summarized into themes based on ongoing discussions and feedback between evaluation team members.

Stakeholder Feedback

At 33 months (July 2012), Action Team leaders were asked to submit brief narrative reports on their team's activities, successes, challenges, and future plans. In total, 8 out of 10 Action Teams submitted reports, reflecting the team leaders' perceptions and opinions; these data provided rich depictions for InspireNet leaders to better understand the internal factors for success within a team. At 36 months (October 2012), a workshop was held to engage key stakeholders in discussions about InspireNet's successes and challenges as they related to network financial sustainability. The workshop included 15 members in person and 4 members participating via WebEx for a total of 19 participants working in small discussion groups. The consequent report is available at [18].

Network Manager/Staff Reports

Annual reports to the funder and to the network's Advisory Council provide a record of activities, member requests, challenges, and the management team's responses to nurture and sustain the network. These reports provide a "behind-the-scenes" account of network activity and provide data on the work completed to support network functions.

Results

Report of Evaluative Data

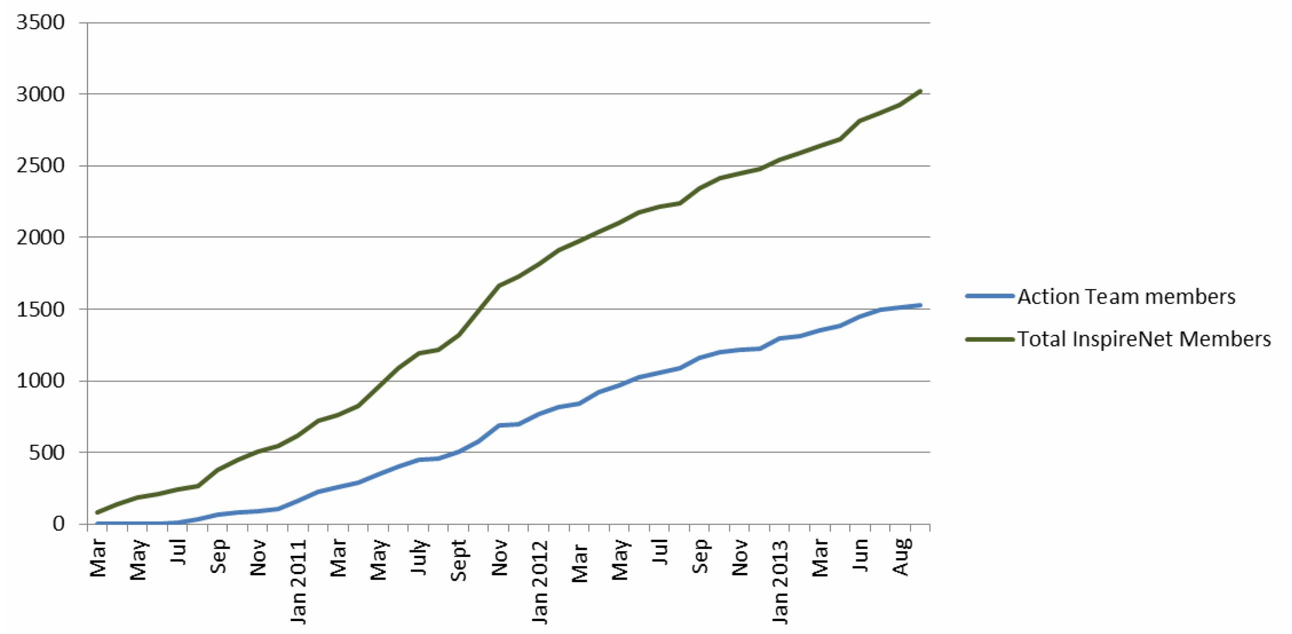
Network Growth

InspireNet's growth continues to increase. At 47 months in the network's life, there are over 3000 members involved to varying degrees with network activities (see Table 1). There are a total of 1442 members participating in Action Teams (see Figure 4). In total, 65.00% (1963/3020) of members work in clinical practice, 25.00% (755/3020) work or study in post-secondary educational institutions, and 10.00% (302/3020) work in smaller organizations and NGOs.

Table 1. InspireNet membership growth.

	12 months, Oct 2010	24 months, Oct 2011	36 months, Oct 2012	47 months, Sep 2013
No. of members	449	1484	2412	3020
% increase	-	231%	63%	25%

Figure 4. Action Team membership in relation to overall membership.



Website Use

Since its launch in May 2010 until September 2013, the website had received more than 76,000 hits with over 370,000 page views by more than 35,000 unique visitors, with an average of 5 pages at over 6 minutes per visit. Highest traffic times occurred immediately following the distribution of InspireNet's biweekly eNews, a digest of all newly curated and hyperlinked website content. The most heavily accessed areas of the website were the teams' eCoPs, followed by InspireNet's "About Us" page, followed by the database. Interestingly, the database had been accessed on average 10 times per day. Of the publically accessible pages, the blogs posts had each been read on average more than 500 times.

Visitors resided primarily in Canada (87.00%, 30,450/35,000) and the United States (5.00%, 1750/35,000). Visitors from other countries resided in the United Kingdom, India, Australia, Philippines, and New Zealand.

Topic-based Action Team eCoPs increased their traffic steadily over time. Since their launch, the number of hits has increased markedly (see Table 2).

Table 2. Cumulative hits for Action Teams' eCoPs.

Action Team[a]	12 months, Oct 2010	24 months, Oct 2011	36 months, Oct 2012	47 months, Sep 2013
eHealth, eTechnologies and Informatics	159	1060	2033	2848
First Nations Health	n/a	16	218	402
Healthy Workplace Climates	90	890	2992	4547
Initiative for a Palliative Approach in Nursing: Evidence & Leadership	n/a	n/a	1214	2017
Interdisciplinary Public Health Club	n/a	n/a	n/a	60
Nurse Educators' Scholarship	206	1839	3687	5047
Nursing Education and Research Rounds	24	1526	3689	5026
Optimal Utilization of Advanced Practice Nursing Roles	262	1485	2041	2100
Practice-based Research Challenge	n/a	93	286	343
Students	n/a	n/a	79	337
Transition of New Grads	n/a	n/a	n/a	531
Total Action Teams' eCoP Hits	741	6909	16,225	23,244

[a] n/a: not applicable where metrics are for a period that is prior to the launch of the team.

Survey Data

Survey #1

The respondent demographics mirrored those of the overall membership: 63.0% (443/703) of the overall membership and 56.0% (70/125) of the respondents worked in clinical practice settings; 29.9% (210/703) of the overall membership and 22.4% (28/125) of the respondents worked in post-secondary educational institutions; and the rest worked in government, NGOs, or other settings. No respondents identified themselves as novice computer users and a majority (69.6%, 87/125) rated themselves as average or proficient computer users. In total, 94.4% (118/125) reported that use of the Internet was useful or very useful to support their work activities, 64.5% (60/93) of the practitioners and managers noted that they had become more interested in embedding nursing health services research evidence in policy and practice as a result of their involvement in InspireNet, and 33.3% (14/42) of practitioners reported that they had changed their practice as a result of learning via the network. Furthermore, 35.7% (15/42) joined one or more of the topic-based Action Teams. Specifically related to network growth, 50.4% (63/125) of respondents reported becoming aware of the network through word of mouth communications and more than 50% (52.0%, 65/125) cited opportunities to network, learn, and collaborate as reasons for becoming a member. Also, 40.8% (51/125) indicated that they were learning new knowledge and skills, primarily related to research through their network membership and 43.2% (54/125) reported that they had or were currently working on research proposals with people they had met through InspireNet. The survey report is available at [19].

Survey #2

The demographic characteristics of the network and the survey respondents remained similar to those in the first survey: 61.0% (128/210) of the survey respondents worked in the health sector, 23.8% (50/210) in post-secondary educational institutions, and the rest in government, NGOs, or other areas. Geographic representation was likewise similar to the first survey response. Most respondents had been network members for one year or more. At this point, 85.2% (179/210) responded that they believed the network added value to provincial research capacity, 81.0% (170/210) responded that InspireNet provided good access to information, and 71.9% (151/210) responded that the network was promoting research in the province. In addition, 18.1% (38/210) of the respondents indicated that they had articles or research papers either published or in the process of development with individuals they had met through the network. Similar to responses from the first survey, a majority (73.8%, 155/210) of the respondents indicated that they had learned about the network through word-of-mouth communications. At this time, survey respondents began commenting on their need for discretionary time to work in a virtual environment and the skills they needed to learn to work virtually. Respondents' written comments described a distinct learning curve with virtual working, particularly use of Web 2.0 interactive technologies. While the majority embraced this learning, some found it challenging. There was a general sense that participants considered virtual working to be the way of the future. The survey report is available at [20]. See Table 3 for a summary of survey activities.

Table 3. Survey data collection.

Survey	Time	Length	No. of reminders	No. of members	No. of respondents	Response rate
1	Spring 2011	4 weeks	3	703	125	17.78%
2	Spring 2012	6 weeks	6	2038	210	10.30%

Member Interviews

In total, 23 members, some of whom were Action Team leaders, were interviewed. The majority of those interviewed were active in InspireNet, being members of 2 or more Action Teams. Interview data provided findings about the network's virtual platform, helping managers to better understand what was of perceived value. Interviews revealed that members benefited from the network in multiple ways, in terms of building research capacity, connecting and communicating with other professionals, and supporting their practice.

The theme of communicating with others to support one's work and learning was prominent throughout the interviews, as illustrated by the following quotes:

> *I have learned a tremendous amount through InspireNet in the area [of] building capacity for health services research, knowledge translation, using an eCoP, and website to support the activities of a research team.*
>
> *[The Action Team] brings everyone to my doorstep.*
>
> *The resources, partnerships, and community of nurse-researchers brought together by InspireNet are incredibly helpful to me in my role supporting nursing research.*
>
> *We need to talk to one another regularly and share resources and talk about how we overcame particular problems with [our] role and just checking in with each other. So [the eCoP has] been an absolutely vital communication link.*

Stakeholder Feedback

Action Team leaders identified the value of webinars to provide educational opportunities, while noting the need for dedicating time to take on facilitation of their teams in their volunteer roles. Stakeholder workshop participants expressed their perceived value in InspireNet's full suite of activities, recommending expanding the scope of the network to beyond the discipline of nursing and exploring ongoing funding options. Last, some stakeholders raised concerns about the usability of the network for users who were not accustomed to working virtually.

Manager/Staff Data

The network management team, in preparing annual reports, evaluated their focus of activities, their requests from the membership, and the problems encountered for each of the three years reported here. The following describes salient areas of growth and learning that emerged from this internal review.

In the first year of the network, the network manager spent considerable time in an educator role to teach members how to work virtually. These health care professionals, while experienced users of computers, did not have experience working in a virtual community, so that lessons, mentoring, and practice with webinars, becoming part of a virtual team, entering into synchronous and asynchronous discussions, and accessing wikis and blogs were necessary. The Action Team leaders needed additional education on leading Web-based discussions, recording them, and teaching others the typical protocols for Web-based conferencing and collaborating.

Further, during the first year, there were issues related to network access and use of public/private websites. One prominent issue was that of firewalls that some new members encountered with their work computers that prohibited access. Another was the need for all members to be educated on copyright protection relating to protected materials they were eager to share with others. Both of these issues were addressed, the first through conversations and actions at institutional levels and the second with a focused educational outreach on copyright legislation in our jurisdiction and directions to help members learn how to access copyrighted articles through the existing provincial health care library system.

During the second year, other issues arose. Some members provided feedback that the website was difficult to navigate. The management team arranged for a faculty member leading a group of students in a University of Victoria School of Health Information Science course to complete a usability evaluation of the website so that improvements could be made. Other issues in the second year were a result of network growth; the need for instruction on working in a virtual environment continued.

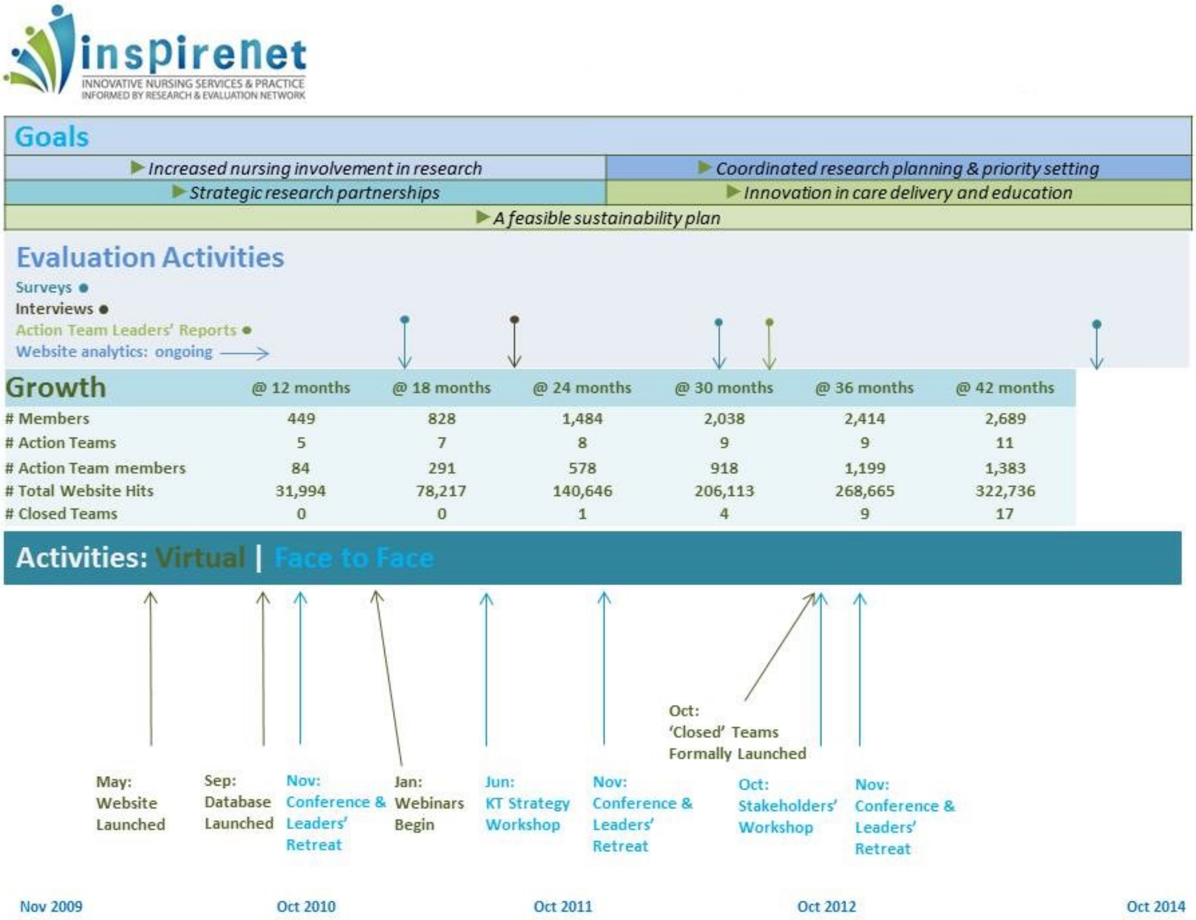
In its third year, the network developed Closed Team eCoPs for established research teams to conduct their work. These team leaders and members required orientation to the network tools and, in some cases, training to adapt them to their own work.

InspireNet has held an annual conference for each of its three years. The conferences have been attended by over 450 people and have provided a venue for continuing education, face-to-face networking, and for members to present their own research to each other. Conference evaluations reported that participants found value in this networking time and appreciated the opportunities to learn from one another. Overwhelmingly, participants reported a high level of satisfaction with these events.

Cumulative Results

Scrutiny of network growth patterns shows us that the network draws new members at the time of the annual conference, that hits on the website coincide with the publication of the biweekly eNewsletter, and that a majority of members entered into the network community after attending or retrieving a recording of a network Action Team webinar. Figure 5 indicates network growth in relation to other network activities, including website traffic, which may be an indicator of active membership.

Figure 5. Timeline of growth and activities, June 2013.



Growth	@ 12 months	@ 18 months	@ 24 months	@ 30 months	@ 36 months	@ 42 months
# Members	449	828	1,484	2,038	2,414	2,689
# Action Teams	5	7	8	9	9	11
# Action Team members	84	291	578	918	1,199	1,383
# Total Website Hits	31,994	78,217	140,646	206,113	268,665	322,736
# Closed Teams	0	0	1	4	9	17

Discussion

Principal Findings

Challenges in Evaluating a Virtual Network

For a network requiring active member participation and largely volunteer leadership, InspireNet has struggled to interpret the low rate of response to its annual evaluation surveys. While conclusions cannot be drawn about the ratio of core members versus peripheral members, network leaders question whether the ratio of core members to peripheral members described in other Web-based groups may be reflected in InspireNet, exhibited by a 10% to 18% response rate to surveys. It cannot be surmised that those who choose not to respond to surveys failed to respond because they are not actively engaged in the network (peripheral), are not getting something out of their network membership, or for some other reason. Further research in the area of active versus peripheral membership in Web 2.0 networks is needed.

Factors for Success in a Virtual Network

InspireNet has connected people across professional roles, academic settings, and geographical boundaries. These connections have fostered the formation of research partnerships and other collaborative work across the province; the technology used by InspireNet has enabled the creation of a network of eCoPs and has been critical to the achievement of the network's goals. Experience of those using eCoPs documents that factors such as voluntary involvement, distributed leadership, shared identity, transparency, accessibility, and being problem-focused have contributed to their success [12]. Further, a recent case study of a professional network for dental care providers describes very similar challenges as those faced by InspireNet [21]. These authors identified the need for a regular schedule of activities in the eCoP, and the need for IT and user support. In addition, these authors acknowledge that while there are obvious benefits of having researcher-practitioner membership in professional eCoPs, there are challenges in terms of making activities relevant to both groups at once. The InspireNet experience is very consistent with these reports and, further, InspireNet has learned that active participation of a network management and staff team serving in a facilitator role is imperative for growth of the network and for welcoming and teaching new network members.

Conclusions and Recommendations

InspireNet's evaluation activities are ongoing, including a social network analysis exercise to better understand how the network operates. Network leaders are encouraged by the strength of membership numbers, which came as a surprise as the network has evolved, and by the outcomes that have been achieved in the first three years of the network. One factor that has proved essential to network growth is the presence of an active, responsive, and supportive network manager and the development of individuals in volunteer leadership roles who meet members' needs to learn how to work in a virtual environment and also serve to welcome each new member. Evaluative data indicates that InspireNet grew because of word-of-mouth communications, that members join to connect with others, and that they increase their contact with the network immediately after network outreach activities. We conclude that InspireNet is meeting members' needs for professional development, social interaction, and support, consistent with what is known about eCoPs and social learning. Because this is a network primarily of nurses, it would be of interest to explore whether this desire for professional social networking is a characteristic of the nursing population and whether or not this characteristic extends to other health care professionals. Further research on this topic is recommended.

Another lesson is the emerging role of "community manager", which is evolving in the field of social media [13]. This role is unique, requiring leaders to facilitate and encourage online participation of community members. Recognizing that a majority of members entered Action Teams through webinars facilitated by team leaders, and remained active because of facilitated discussions on topics of interest, it is believed that having individuals step into this role as network Action Team leaders is a necessary part of both attracting and sustaining members. Last, InspireNet's experience indicates that development of network tools alone is not sufficient for the growth and success of a professional network. A network team that communicates regularly with members as well as providing education, support, and evaluation of network activities is necessary to grow and support a network.

Acknowledgments

InspireNet acknowledges with appreciation funding of the network in full by the BC Nursing Research Initiative through the Michael Smith Foundation for Health Research for the period Nov 2009-Oct 2014.

Conflicts of Interest

None declared.

Multimedia Appendix 1

InspireNet's website details.

[PDF File (Adobe PDF File), 6KB - jmir_v16i2e49_app1.pdf]

References

1. Frisch N, Borycki EM, Mickelson G, Atherton P, Novak-Lauscher H, Hooker D, et al. Use of social media and Web 2.0 technologies to increase knowledge and skills of British Columbia nurses. 2012 Jun 01 Presented at: 11th International Congress on Nursing Informatics; 2012; Montreal, Canada URL: http://knowledge.amia.org/amia-55142-cni2012-1.129368/cni2012-1.130355/f-001-1.130356/a-038-1.131723/an-038-1.131725?qr=1
2. Bassell K. Social media and the implications for nursing faculty mentoring: A review of the literature. Teaching and Learning in Nursing 2010 Oct;5(4):143-148 [FREE Full text] [doi: 10.1016/j.teln.2010.07.007]
3. Bristol TJ. Twitter: consider the possibilities for continuing nursing education. J Contin Educ Nurs 2010 May;41(5):199-200. [doi: 10.3928/00220124-20100423-09] [Medline: 20481418]
4. Peate I. The community nurse and the use of social media. Br J Community Nurs 2013 Apr;18(4):180-185. [Medline: 23574909]
5. Fraser R. Social media: An innovative way to share nursing ideas. Reflections on Nursing Leadership. In: Honor Society of Nursing, Sigma Theta Tau International. Indianapolis, IN: Sigma Theta Tau International; 2010.
6. Fogelson NS, Rubin ZA, Ault KA. Beyond likes and tweets: an in-depth look at the physician social media landscape. Clin Obstet Gynecol 2013 Sep;56(3):495-508. [doi: 10.1097/GRF.0b013e31829e7638] [Medline: 23835911]
7. Bell JM. Social media and family nursing: where is my tribe? J Fam Nurs 2010 Aug;16(3):251-255 [FREE Full text] [doi: 10.1177/1074840710378404] [Medline: 20686101]
8. Winterfeldt LG, Fox LM. Medical Library Association News. 2009 May. Technology: The next wave in the Web 2.0 revolution: Professional Networks URL: http://www.mlanet.org/publications/mlanews/ [accessed 2014-01-29] [WebCite Cache ID 6Mzxc8BXl]
9. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
10. Li LC, Grimshaw JM, Nielsen C, Judd M, Coyte PC, Graham ID. Evolution of Wenger's concept of community of practice. Implement Sci 2009;4:11 [FREE Full text] [doi: 10.1186/1748-5908-4-11] [Medline: 19250556]
11. Wenger E, McDermott RA, Snyder WM. Cultivating communities of practice: a guide to managing knowledge. Boston, Mass: Harvard Business School Press; 2002.

12. Ho K, Jarvis-Selinger S, Norman CD, Li LC, Olatunbosun T, Cressman C, et al. Electronic communities of practice: guidelines from a project. J Contin Educ Health Prof 2010;30(2):139-143. [doi: 10.1002/chp.20071] [Medline: 20564704]
13. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]
14. Lave, Wenger E. Situated learning: legitimate peripheral participation. Cambridge [England]: Cambridge University Press; 1991.
15. MacDonald J, Atkin W, Daugherity F, Fox H, MacGillivray A, Reeves-Lipscomb D, et al. CPsquare Foundations of Communities of Practice Workshop. 2003. Let's get more positive about the term 'lurker' URL: http://cpsquare.org/2003/08/lets-get-more-positive-about-the-term-lurker/ [accessed 2013-10-10] [WebCite Cache ID 6KGwksfOh]
16. InspireNet. Innovative Nursing Services and Practice Informed by Research and Evaluation Network URL: http://www.inspirenet.ca/ [accessed 2014-01-31] [WebCite Cache ID 6N2VGgZiz]
17. Lasker R, Weiss E, Miller R. Partnership synergy: a practical framework for studying and strengthening the collaborative advantage. Milbank Q 2001;79(2):179-205, III [FREE Full text] [Medline: 11439464]
18. InspireNet. 2012. Workshop Report URL: http://www.inspirenet.ca/Sustainability-Team [accessed 2014-01-31] [WebCite Cache ID 6N2VQiw8S]
19. InspireNet. 2011. Survey #1 Report URL: http://www.inspirenet.ca/document/inspirenet-members-survey-report-march-2011 [accessed 2014-01-31] [WebCite Cache ID 6N2VZmlUP]
20. InspireNet. 2012. Survey #2 Report URL: http://www.inspirenet.ca/document/inspirenet-members-survey-2012 [accessed 2014-01-31] [WebCite Cache ID 6N2ViWrFV]
21. Friberger MG, Falkman G. Collaboration processes, outcomes, challenges and enablers of distributed clinical communities of practice. Behaviour & Information Technology 2013 Jun;32(6):519-531. [doi: 10.1080/0144929X.2011.602426]

Abbreviations

eCoP: electronic communities of practice
FTE: full-time equivalent
InspireNet: Innovative Nursing Services and Practice Informed by Research and Evaluation Network

Edited by G Eysenbach; submitted 16.10.13; peer-reviewed by J Brixey, R Fraser, M Malliarou; comments to author 15.11.13; revised version received 28.11.13; accepted 03.12.13; published 21.02.14.

Please cite as:
Frisch N, Atherton P, Borycki E, Mickelson G, Cordeiro J, Novak Lauscher H, Black A
Growing a Professional Network to Over 3000 Members in Less Than 4 Years: Evaluation of InspireNet, British Columbia's Virtual Nursing Health Services Research Network
J Med Internet Res 2014;16(2):e49
URL: http://www.jmir.org/2014/2/e49/
doi:10.2196/jmir.3018
PMID:24566806

Original Paper

Are Public Health Organizations Tweeting to the Choir? Understanding Local Health Department Twitter Followership

Jenine K Harris[1], PhD; Bechara Choucair[2,3], MD; Ryan C Maier[4], BA; Nina Jolani[5], BA; Jay M Bernhardt[6], MPH, PhD

[1]Brown School, Washington University in St Louis, St Louis, MO, United States
[2]Chicago Department of Public Health, Chicago, IL, United States
[3]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[4]Center for Public Health Systems Science, Washington University in St Louis, St Louis, MO, United States
[5]National Association of County and City Health Officials, Washington DC, DC, United States
[6]Center for Digital Health and Wellness, College of Health and Human Performance, University of Florida, Gainesville, FL, United States

Corresponding Author:
Jenine K Harris, PhD
Brown School
Washington University in St Louis
One Brookings Drive
Campus Box 1196
St Louis, MO, 63130
United States
Phone: 1 3149353522
Fax: 1 3149353756
Email: harrisj@wustl.edu

Abstract

Background: One of the essential services provided by the US local health departments is informing and educating constituents about health. Communication with constituents about public health issues and health risks is among the standards required of local health departments for accreditation. Past research found that only 61% of local health departments met standards for informing and educating constituents, suggesting a considerable gap between current practices and best practice.

Objective: Social media platforms, such as Twitter, may aid local health departments in informing and educating their constituents by reaching large numbers of people with real-time messages at relatively low cost. Little is known about the followers of local health departments on Twitter. The aim of this study was to examine characteristics of local health department Twitter followers and the relationship between local health department characteristics and follower characteristics.

Methods: In 2013, we collected (using NodeXL) and analyzed a sample of 4779 Twitter followers from 59 randomly selected local health departments in the United States with Twitter accounts. We coded each Twitter follower for type (individual, organization), location, health focus, and industry (eg, media, government). Local health department characteristics were adopted from the 2010 National Association of City and County Health Officials Profile Study data.

Results: Local health department Twitter accounts were followed by more organizations than individual users. Organizations tended to be health-focused, located outside the state from the local health department being followed, and from the education, government, and non-profit sectors. Individuals were likely to be local and not health-focused. Having a public information officer on staff, serving a larger population, and “tweeting” more frequently were associated with having a higher percentage of local followers.

Conclusions: Social media has the potential to reach a wide and diverse audience. Understanding audience characteristics can help public health organizations use this new tool more effectively by tailoring tweet content and dissemination strategies for their audience.

(J Med Internet Res 2014;16(2):e31) doi:10.2196/jmir.2972

KEYWORDS

local health department; Twitter; social media

Introduction

In the United States, local health departments are governmental agencies providing essential public health services in cities, counties, metropolitan areas, districts, and tribal areas [1]. One of the services provided by local health departments throughout the United States is informing and educating constituents about health [2]. Communication with constituents about public health issues and health risks is also among the recently developed standards required of local health departments for accreditation [3]. Past research found that only 61% of local health departments met standards for informing and educating constituents [4], suggesting a considerable gap between current practices and best practice. In addition, the most recent data from the national professional organization representing local health departments, the National Association of County and City Health Officials (NACCHO), shows nearly half of local health departments (45.5%) received budget cuts in 2010 [5], indicating a growing need for local health departments to implement low-cost strategies when providing essential services in difficult economic times.

Approximately 80% of US adult Internet users have searched for health information online, making the Internet second only to health care providers as a source for health information [6]. Social media, such as Twitter, Facebook, and YouTube, have emerged as extremely popular online platforms for health information seeking and information sharing [6-8]. Social media platforms are increasingly used by health care providers [7,9-11] and public health practitioners [12-21] to find and share health information, conduct surveillance, and manage large-scale emergency situations. Twitter is a popular free-to-use social media application for microblogging, or sending and receiving brief, direct, one-to-many messages or “tweets” [12,22]. Twitter accounts can be followed by other Twitter users, allowing individuals or organizations to receive information and share or retweet these messages to others in their network. Retweeting is forwarding a tweet sent by another user, usually adding “RT” to the text to show the tweet is not original. As of 2013, approximately 18% of US adults reported using Twitter [23]. Twitter use among US adults is associated with younger age and is higher among black non-Hispanic Internet users. However, Twitter use is independent of gender, educational attainment, and income [24,25], suggesting it may provide an important new channel for disseminating public health messages to groups, such as lower socioeconomic status, households that have traditionally been more difficult to reach with health information.

Understanding the audience intentionally receiving health information is key to successful health communication [26-29]. We know the general characteristics of Twitter users overall; however, little is known about the composition of social media audiences for specific types of Twitter users, such as local health departments [24,30,31]. To aid in developing the evidence base for local health department social media use, a promising low-cost strategy for educating and informing constituents about health, we examined: (1) the general characteristics of local health department Twitter followers, such as whether they are individuals or organizations; (2) the relationship between health department characteristics, such as size and staffing, and Twitter follower characteristics; and (3) the relationship between local health department Twitter use, such as tweet frequency and Twitter follower characteristics.

Methods

Overview

As of July 30, 2012, we identified 217 local health departments nationwide using Twitter (identification process described elsewhere [32]) out of more than 2500 local health departments in the United States. At a minimum, Twitter users choose a username when opening an account. They may also enter a photo, their full name, location, description, and affiliated website link, although this information is not required for an account. In 2013, we used NodeXL [33], an open-source network data collection tool, to collect the Twitter followers for each of the 217 local health departments using Twitter. All 146,013 followers were consolidated into a single data file comprised of more than 98,000 unique Twitter users. To better understand the composition of this large group of followers, we constructed a representative stratified random sample of approximately 5000 Twitter users for in-depth coding. To ensure local health departments with fewer followers were represented in the dataset, we first selected a random sample of 59 of the 217 local health departments. We then compared the local health departments in the sample to those not in the sample to determine whether the sampled health departments were representative of the population of local health departments using Twitter. The comparison included 3 characteristics associated with Twitter adoption and Twitter followership for a local health department: jurisdiction population, spending per capita, and public information officer staffing [32]. We found no significant differences between those selected and those not selected for jurisdiction population size (t_{199}=0.60; P=.55), spending per capita (t_{166}=1.34; P=.18), and employment of a public information officer (χ^2_1=0.4; P=.51). We took a random sample of 100 followers from each of the 59 local health departments. In all, 21 of 59 (36%) local health departments had fewer than 100 followers, so these local health departments contributed all of their followers to the sample for coding, resulting in a sample of 4779 Twitter followers from 59 local health departments.

Coding

Textbox 1. Coding scheme for Twitter followers (N=4779) from 59 local health departments nationwide.

Location: Where is the Twitter follower located?

- In-state where local health department is located
- Out-of-state (in United States)
- Outside United States
- Unable to determine (in United States)
- Unable to determine

Type: Based on the username and description, does the follower account appear to be for an...?

- Individual
- Organization/business
- Unable to determine

Industry: Which of the following industries best describes the follower based on description or linked website?

- Spam
- Unable to determine
- Health-focused
  - Private user
  - Educational institution
  - US government
  - Local
  - State
  - National
  - Campaign/program
  - Not-for-profit (nongovernmental)
  - Local
  - State
  - National
  - International
  - Campaign/program
  - For profit
  - Hospital or hospital system
  - Private physician or physician offices
  - Drug company or representative
  - Managed care
  - Patient advocate
  - Medical device maker, seller
  - Fitness center/gym/personal training
  - Diet/nutrition (eg, Weight Watchers, Jenny Craig)
  - Assisted living
  - Media
  - TV
  - Newspaper/magazine (print media)
  - Radio
  - Social media or website (not affiliated with TV, radio, newspaper)

  - Public relations firm
  - Other (make note)
- Not health-focused
  - Private user
  - Educational institution
  - US government
  - Local
  - State
  - National
  - Campaign/program
  - Not-for-profit (nongovernmental)
  - Local
  - State
  - National
  - International
  - Campaign/program
  - For profit
  - Media
  - TV
  - Newspaper/magazine (print media)
  - Radio
  - Social media or website (not affiliated with TV, radio, newspaper)
  - Public relations firm
  - Other (make note)

Project team members reviewed the Twitter follower information to develop a coding scheme with 3 broad categories: location, type, and industry. Location was an indicator of whether the Twitter follower was in the same state with the local health department, out-of-state (in the United States), outside the United States, unable to tell but within the United States, and unable to determine. Type of follower included 3 categories: individual, organization/business, and unable to determine. To discern individual followers, we looked for information written in the first person using “I” or “my” or other descriptors or wording indicating the user was an individual. Organizational accounts, on the other hand, often included a statement of organizational purpose. Industry was divided into 4 subgroups: health-related, non-health-related, spam, and unable to determine. Classification of an account as spam occurred when the account did not appear to be a legitimate person or business. For example, one spam account was following 2002 others, had 67 followers, and had never tweeted. Another spam-classified account had no user description, 49 followers, and had tweeted over 1000 times with nearly all the tweets being retweeted from another user, suggesting automated retweeting by a spambot, or a program that is designed to send out spam. Within industry were several specific types of organizations. If the industry was not easy to glean from the Twitter profile (eg, “St. Mary’s is a nonprofit hospital”), we searched the user’s website and often found the type on the About Us page. Consistent with previous research [34], industry was classified for both organizations and individual users whose Twitter profiles indicated they were representing an organization. Textbox 1 includes a summary of the coding scheme.

To test the reliability of the coding scheme, 4 coders coded data from the same 100 Twitter followers (2.09%) randomly selected from the overall sample of 4779. Krippendorff’s alpha for nominal data was computed for each of the 3 broad coding categories (type, location, industry). For follower type, alpha was .70 (95% CI .63-.77), for follower location alpha was .88 (95% CI .84-.91), and for industry alpha was .68 (95% CI .64-.72). Given these acceptable reliability scores, the full dataset was divided and coded independently by the 4 coders.

Analyses

For the first aim of the study, understanding the general characteristics of local health department Twitter followers, frequencies and percentages were used to examine the distribution of follower types. Chi-square tests were used to determine whether certain types of followers were more or less likely than expected given the overall distribution of followers. For example, were individual users more or less likely than expected to be local? Standardized residuals were calculated for significant chi-square test results; standardized residuals greater than 1.96 indicated significantly more followers than expected fell into a given category, whereas standardized residuals less than –1.96 indicated significantly fewer followers than expected fell into a category. Followers classified as “unable to determine” were omitted from analyses.

The second and third aims were to understand the relationship between local health department characteristics, Twitter use, and characteristics of their Twitter followers. For these aims, follower data was aggregated by local health department to compare the proportions of follower types in local health departments varying by (1) Twitter use (number of followers, number of tweets) and (2) resources (population size, staffing, funding per capita). Twitter usage was obtained through NodeXL in April 2013. Local health department resource information was obtained from the NACCHO 2010 Profile Study.

Because local health departments provide services to their local constituents, it is important to know what is associated with reaching local individuals. Local health departments with a public information officer may have more, and more organized, information-sharing efforts in the local community given this specialized staffing. Larger jurisdiction population and a higher number of tweets are associated with more followers [32], but it is not known whether these factors also influence the proportion of local followers. To examine what is associated with reaching local individuals, we hypothesized that:

1. Local health departments with a public information officer have a higher proportion of local followers and individual followers than local health departments without a public information officer.
2. The more Twitter followers and tweets a local health department has, the higher the percentages of local and individual Twitter followers there will be.
3. The larger the jurisdiction population, the higher the percentages of local Twitter followers and individual Twitter followers there will be.

Other ways for local health departments to reach and inform local constituents could be through local media and local government. Journalists have adopted social media as sources of information, with more than 30% of print journalists deeming social media as important or very important as of 2009 [35]. Local media following local health departments on Twitter may facilitate secondary dissemination by covering tweeted topics in local newspapers or on local radio or television broadcasts. Media advocacy is one strategy that may work to influence policy [36]; local government following local health departments on Twitter could use tweeted information to support local policy development, passage, or enforcement. To understand more about connections with local media and government, we hypothesized:

1. Local health departments with a public information officer have a higher proportion of local media followers (TV, radio, and newspaper) and local government followers than local health departments without a public information officer.
2. The larger the jurisdiction population, the higher the percentages of local media (TV, radio, and newspaper) followers and local government followers there will be.

Hypotheses 1, 3, 4, and 5 aid in addressing aim 2 (understanding the relationship between local health department characteristics and Twitter use), whereas hypothesis 2 aids in addressing aim 3 (understanding the relationship between local health department characteristics and characteristics of their Twitter followers).

Results

Summary

The 59 local health departments in the sample had between 9700 and 3 million constituents in their local jurisdictions according to the 2010 NACCHO Profile Study. In all, 29 (49%) of the departments reported having a public information officer, whereas 23 (39%) reported not having one (7/59, 12% were missing data on this staffing). The median number of Twitter followers was 218 (range 7-11,827; mean 770.1, SD 1688.0); the median number of sent tweets per health department since adopting Twitter was 324 (range 0-5849; mean 667.9, SD 1083.1). Health departments in the sample joined Twitter between June 2008 and January 2012; more than half (34/59, 58%) joined in 2009.

Twitter Follower Characteristics

Overall, we found that local health departments had more Twitter followers that were organizations (2591/4434, 58.43%) than individuals (1843/4434, 41.57%). Of the 1843 individuals, 1267 (68.75%) were private personal accounts (private users not affiliated with a specific organization or business). The 1267 private individuals comprised 29.07% of the follower industry classifications for the 4359 classified by industry (Table 1). The most common type of organizational Twitter follower was from the for-profit sector (n=1053), comprising 24.16% of the 4359 followers classified by industry. More than half of followers classified as health-focused (2592/4340, 59.72%) did not include a health focus in their account information and 2149 of 3878 location-classified followers (55.42%) were in the same state as the health department they were following. A summary of the type, industry, and location of local health department Twitter followers is shown in Table 1. Users classified as “unable to determine” during coding were not included in analyses or shown in Table 1. Overall, those classified as unable to determine comprised 7.22% of the type category (345/4779), 8.79% of the industry category (420/4779), 9.19% of the health-focus category (439/4779), and 18.85% of the location category (901/4779).

Omitting the private user category because it only applied to individuals, follower industry was significantly associated with follower type (χ^2_6=112.6, P<.001). Standardized residuals indicated that there were more organizations and fewer individuals than expected in the nonprofit category (Table 1). The opposite was true for the media category with significantly more individuals than expected and fewer organizations than expected. In addition, fewer individuals than expected fell into the other category for industry. Examples of the descriptions entered by users for several common categories of local health department Twitter followers are shown in Table 2.

Considering all cases (including private users), industry was associated with health focus ($\chi^2_7=783.3$, *P*<.001). The proportion of health-focused followers was higher than expected for education, government, nonprofit, and other. The proportion of health-focused followers was lower than expected in the private user, for-profit, and media categories. There was a significant association between type and health focus ($\chi^2_2=308.3$, *P*<.001) with more organizations and fewer individuals being health-focused than expected. Type and location were also significantly associated ($\chi^2_3=47.0$, *P*<.001), with more organizations than expected being outside the state from the local health department and more individuals than expected being within the same state as the local health department.

Table 1. Characteristics of local health department Twitter followers.

Characteristic	All		Individuals		Organizations		*P*
	n	%	n	%	n	%	
Type	4434						
Individual	1843	41.57	—	—	—	—	
Organization	2591	58.43	—	—	—	—	
Industry	4359		477		1551		<.001[a]
Private user	1267	29.07	—	—	—	—	
Education	160	3.67	24	5.03	136	5.33	
Government	497	11.40	71	14.88	426	16.70	
Nonprofit	556	12.76	31	6.50[b]	525	20.58[c]	
For-profit	1053	24.16	174	36.48	868	34.03	
Media	599	13.74	160	33.54[c]	422	16.54[b]	
Other	192	4.40	16	3.35[b]	170	6.67	
Spam	35	0.80	1	0.21	4	0.16	
Health focus	4340		1738		2563		<.001
Yes	1748	40.28	419	24.11[b]	1303	50.83[c]	
No	2592	59.72	1319	75.89[c]	1260	49.16[b]	
Location	3878		1371		2417		<.001
In-state	2149	55.42	840	61.27[c]	1281	53.00[b]	
Out-of-state	1195	30.81	363	26.48[b]	815	33.72[c]	
In US state unknown	337	8.69	41	2.99[b]	147	6.08[b]	
Outside US	197	5.08	127	9.26	174	7.20	

[a] Private person was omitted for the purposes of bivariate analysis.

[b] More followers than expected fell into this category (standardized residuals >1.96).

[c] Fewer followers than expected fell into this category (standardized residuals <–1.96).

Table 2. Examples of common categories of local health department Twitter followers.

Follower type	Health focus	Twitter user description
Private person	No	I love music and travel. Watching movies and sunsets. I like to play World of Warcraft.
	Yes	Huggable Health Educator
Nonprofit organizations	No	IBA is a nonprofit agency dedicated to empowering individuals and families through education, economic development, technology and the arts.
	Yes	Advocating for the health and dignity of Denver's injection drug users in accordance with #harmreduction principles. Syringe exchange and Naloxone Save Lives!
Media individuals	No	San Diego reporter at The Daily Transcript. Freelance sports writer for Southwest Riverside News Network. Associate Producer for KUSI Prep Pigskin Report.
	Yes	Journalist covering medical/health & fitness and writing features for the Tyler Morning Telegraph.

Local Health Department Characteristics

On average, more than half of a local health department's Twitter followers were from within the state (5%, range 10%-91%). Consistent with the overall composition of the follower sample, 70% (41/59) of local health departments were followed by a higher proportion of organizations than individuals (range 22%-83% organizational followers). An even greater number of departments had a majority of non–health-focused followers (compared to health) with 44 of 59 (75%) departments having more than 50% non–health-focused followers (range 17%-86%). The average percentage of in-state media followers (TV, newspaper, radio) was 6% (SD 5.8) with a range from 0 to 25%. The average percentage of in-state government followers was 7% (SD 8.1) with a range from 0 to 43%.

Local and Individual Followers

We hypothesized that the local health departments with a public information officer have a higher proportion of local followers and individual followers than local health departments without a public information officer. A *t* test indicated that local health departments with a public information officer did have a significantly higher percentage of local followers (t_{50}=2.3, *P*=.03) than local health departments without a public information officer. Local health departments with a public information officer had an average of 63.9% (range 15.1%-90.8%) in-state followers compared to 49.3% (range 9.5%-90.9%) for those without a public information specialist. However, another *t* test ($t_{35.9}$=–1.71; *P*=.10) found no significant difference in the proportion of followers that were individuals among local health departments with a public information officer (mean 44.1%, range 28.6%-55.9%) compared to those without a public information officer (mean 38.9%, range 21.7%-66.7%).

We hypothesized that, the more Twitter followers and tweets a local health department had, the higher the percentages of local and individual Twitter followers there was. There was a strong positive correlation between the number of followers and the number of tweets for a local health department (*r*=0.74, *P*=.049), so we tested only the relationship between number of tweets and follower characteristics. We chose number of tweets because it is a characteristic the local health department can modify as part of a social media strategy, unlike the number of followers. We found a weak, positive, and statistically significant association between the number of tweets a health department had sent and the proportion of its followers classified as individuals (*r*=.32, *P*=.02); however, there was no significant association between number of tweets and the proportion of in-state followers.

We hypothesized that, the larger the jurisdiction population, the higher the percentages of local Twitter followers and individual Twitter followers. A correlation coefficient indicated a positive and significant, although weak, association between jurisdiction population and percent of followers who were local (*r*=0.33, *P*=.01). However, population size had no significant association with proportion of followers who were individuals.

Local Media and Local Government Followers

We also hypothesized that local health departments with a public information officer have a higher proportion of local media followers (TV, radio, and newspaper) and local government followers than local health departments without a public information officer. The *t* tests demonstrated no significant association between having a public information officer and having more local media or local government followers.

Finally, we hypothesized that, the larger the jurisdiction population, the higher the percentages of local media (TV, radio, and newspaper) followers and local government followers. The size of the jurisdiction population was not associated with the percent of followers who were local government or local media.

Discussion

Principal Findings

The results of our research reveal that local health departments on Twitter in the United States are followed by more organizations than individuals. Many of the organizations are health-focused, out-of-state, and from the education, government, and nonprofit sectors, suggesting that there may be a communication network comprised of organizations with a health-related mission developing on Twitter. Individual followers of local health departments, on the other hand, tended to be locally based and largely did not have a health focus, indicating that, at least where individuals are concerned, local health departments may not be just tweeting to members of the public who are already health-focused (ie, the choir).

Evidence of a network of public health organizations developing on Twitter is consistent with at least 2 recent studies [37,38]. The first study demonstrated that many of the local health departments on Twitter follow the same set of government agencies, media, nonprofit organizations, professional associations, educational institutions, and for-profit organizations, many of which are health-focused [38]. Several organizations were followed by 100 or more local health departments, including 4 Centers for Disease Control accounts (@CDCemergency, @CDCFlu, @CDC_eHealth, @CDCgov), 2 other government agencies (@FDArecalls, @HHSGov), and 2 professional organizations (@PublicHealth, @NACCHOalerts). The second study identified a 1516 Twitter follower relationship among 182 local health departments on Twitter. Contrary to the current study, however, the follower connections among these 182 local health departments were more likely to be between organizations in the same state than to be across states [37].

Consistent with other studies [12,32], more tweeting by local health departments was associated with having more followers and, in this case, with having more individual followers (as opposed to organizational). A recent analysis of local health department tweet content indicated that local health departments are largely tweeting about healthy behaviors, most likely with the purpose of reaching individual constituents [20]. Having more individual followers appears consistent with sending these individually focused tweets. However, because more than half the Twitter audience is comprised of organizations, and growing evidence around interorganizational connections for public health on Twitter, the tweet content study [20] suggests a possible disconnect with individually focused tweets reaching a Twitter audience comprised of organizations.

Although tweeting more often and serving larger populations has been associated with having more followers overall for local health departments [39], we found these characteristics were not associated with having more local followers in general or more local media and local government followers specifically. Past research has also demonstrated that employment of a public information officer in a local health department is associated with adopting social media earlier, having more followers [39], and tweeting about specific public health topics [40]. Although local followership was significantly higher for health departments employing a public information officer in this study, having a public information officer was not associated with having more local media or government followers. Given the emerging evidence about social media activity in local health departments employing public information specialists, additional research is needed into the social media strategies of these specialized staff members and other characteristics of their health departments that might influence social media activity.

Our findings may be useful for local health departments in at least two ways. First, we identified the characteristics (eg, jurisdiction population size, employing a public information officer) and practices (eg, tweet frequency) associated with local health department Twitter follower characteristics. This information could inform strategic planning for local health departments using or considering using Twitter. For example, if a goal of a local health department is to reach greater numbers of individuals rather than organizations, their Twitter strategy could include a regular daily or weekly tweeting schedule.

Second, understanding who the Twitter followers are could help local health departments better target tweets to diverse audiences. For example, local media and policymakers may be important followers for a local health department. Standard strategies (eg, tweeting more, developing an easy-to-find user profile [41]) for increasing the number of followers may bring in more media and policymakers. However, additional research into how media and policymakers on Twitter select specific information sources may help to identify strategies for local health departments in increasing the presence of these followers. As another example, out-of-state organizations are less likely to be interested in tweets focused at the individual, as well as locally focused or locally relevant tweets about health and health-related events. However, these followers may be interested in learning about innovative or successful local health department programs and best practices. Local health departments wishing to connect and engage more with this existing audience might program their Twitter accounts as dissemination platforms to reach these organizations with relevant information about successful strategies and programs.

Some local health departments are already focused on reaching specific individual and organizational audiences. For example, the Chicago Department of Public Health has begun conducting campaigns using social media as a dissemination channel and making explicit efforts to interact with local individuals through Twitter activities, such as Twitter live chats. One of these events took place in early 2013, when Dr Julie Morita of the Chicago Department of Public Health answered questions from Chicagoans in a Twitter live chat about flu (using hashtag #FluChicago) just as local news coverage of flu was increasing. Through their official Twitter account (@ChiPublicHealth), she answered questions ranging from the Chicago mayor asking about prevention when in close contact with many people (Figure 1) to constituents asking about the severity of the flu season, the safety and location of flu shots, and how long to wait before returning to work after having the flu. Although @ChiPublicHealth chose to use their Twitter feed to combat an emerging local public health problem, a recent study found examining diabetes rates and tweeting about diabetes found no association between local rates and local health department tweeting [40], suggesting that opportunities exist to increase Twitter use to address locally relevant health issues.

Other health departments have adopted Twitter with the purpose of sharing information with public health organizations. For example, local health departments across Utah made a statewide effort to adopt and use social media. A local health department practitioner at Bear River Health (@BearRiverHealth) in Utah described this strategy as follows, “Not only does it allow us an opportunity to share information, it allows us to communicate in a new way with the communities that we serve together as a state. For example, when we launch an immunization campaign we now have the ability to share the same message seamlessly across our entire state through Twitter and Facebook. We share one another's posts, comment on status, and generally connect" (Jill Parker, personal communication, November 2012). This coordinated effort and active use of Twitter has resulted in Utah communities such as the jurisdiction of Bear River Health with a 2010 population of 163,836 to reach more than 3200 followers, more than 5 times the average number of followers for a local health department [32]. Given these innovative and varied uses of Twitter by local health departments and the composition of follower types for local health departments, future studies may wish to focus on whether and how local health departments are choosing their social media strategies and which strategies are effective at influencing follower behavior for different types of followers.

Figure 1. Chicago mayor Rahm Emanual participating in the #FluChicago 2013 Twitter chat about flu prevention with @ChiPublicHealth during flu season (Photo from Chicago Mayor's Office, @ChicagosMayor [43]).

Limitations and Conclusions

Limitations to this study include cross-sectional data, reliance on self-reported information, and a lack of information on follower engagement. For example, because many Twitter users do not include geographically specific information about their location in the user profile, the coding of local was limited to identifying whether a follower was in the same state as the local health department. Likewise, a Twitter user may have had a health focus, but did not include health-related language in their profile and was coded as nonhealth. In addition, without additional information about Twitter follower engagement, such as mentions and retweets, it is impossible to know the extent to which the followers were actively engaged with the health departments through their Twitter accounts [42]. Despite its limitations, this study provides an important first look at the characteristics of Twitter users connected to local health departments.

Acknowledgments

Support for this research was provided by a Mentored Research Scientist Development Award from the Robert Wood Johnson Foundation through the National Coordinating Center for Public Health Services and Systems Research.

Conflicts of Interest

None declared.

References

1. National Association of County and City Health Officials. 2005. Operational definition of a functional local health department URL: http://www.naccho.org/topics/infrastructure/accreditation/OpDef.cfm [accessed 2013-11-16] [WebCite Cache ID 6LAFG8Ndc]
2. Centers for Disease Control and Prevention, National Public Health Performance Standards Program (NPHPSP). 2010 Jul 09. 10 Essential Public Health Services URL: http://www.cdc.gov/nphpsp/essentialservices.html [accessed 2013-09-23] [WebCite Cache ID 6JqfMLG6D]
3. Public Health Accreditation Board Standards and Measures. Alexandria, VA: Public Health Accreditation Board; 2011. URL: http://www.phaboard.org/wp-content/uploads/PHAB-Standards-and-Measures-Version-1.0.pdf [accessed 2013-11-16] [WebCite Cache ID 6LAqTGqLC]
4. Mays GP, McHugh MC, Shim K, Lenaway D, Halverson PK, Moonesinghe R, et al. Getting what you pay for: public health spending and the performance of essential public health services. J Public Health Manag Pract 2004 Oct;10(5):435-443. [Medline: 15552769]
5. 2010 National Profile of Local Health Departments. Washington, DC: National Association of County and City Health Officials; 2011. URL: http://www.naccho.org/topics/infrastructure/profile/resources/2010report/upload/2010_Profile_main_report-web.pdf [accessed 2013-09-23] [WebCite Cache ID 6JqfSMfzc]
6. Fox S. The social life of health information, 2011. Washington, DC: Pew Internet & American Life Project; 2011 May 12. URL: http://pewinternet.org/~/media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2013-09-23] [WebCite Cache ID 6JqfU0Mhf]
7. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
8. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
9. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009;28(2):361-368 [FREE Full text] [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
10. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
11. Kostkova P, Fowler D, Wiseman S, Weinberg JR. Major infection events over 5 years: how is media coverage influencing online information needs of health care professionals and the public? J Med Internet Res 2013;15(7):e107 [FREE Full text] [doi: 10.2196/jmir.2146] [Medline: 23856364]
12. Schein R, Kumanan W, Keelan J. Literature review on effectiveness of the use of social media: a report for Peel public health. 2011. URL: http://www.peelregion.ca/health/resources/pdf/socialmedia.pdf [accessed 2013-09-23] [WebCite Cache ID 6JqfVr1OJ]
13. Jones B. Mixed uptake of social media among public health specialists. Bull World Health Organ 2011 Nov 1;89(11):784-785 [FREE Full text] [doi: 10.2471/BLT.11.031111] [Medline: 22084522]
14. Sublet V, Spring C, Howard J, National Institute for Occupational SafetyHealth. Does social media improve communication? Evaluating the NIOSH science blog. Am J Ind Med 2011 May;54(5):384-394. [doi: 10.1002/ajim.20921] [Medline: 21246585]
15. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
16. Hagar C. Crisis informatics: Perspectives of trust – is social media a mixed blessing? Student Research Journal 2013;2(2):2 [FREE Full text]
17. Yates D, Paquette S. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. International Journal of Information Management 2011 Feb;31(1):6-13. [doi: 10.1016/j.ijinfomgt.2010.10.001]
18. Hingle M, Yoon D, Fowler J, Kobourov S, Schneider ML, Falk D, et al. Collection and visualization of dietary behavior and reasons for eating using Twitter. J Med Internet Res 2013;15(6):e125 [FREE Full text] [doi: 10.2196/jmir.2613] [Medline: 23796439]
19. 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]
20. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of twitter among local health departments: an analysis of information sharing, engagement, and action. J Med Internet Res 2013;15(8):e177 [FREE Full text] [doi: 10.2196/jmir.2775] [Medline: 23958635]

21. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
22. The health communicator's social media toolkit. Atlanta, GA: Centers for Disease Control and Prevention; 2011. URL: http://www.cdc.gov/socialmedia/tools/guidelines/pdf/socialmediatoolkit_bm.pdf [accessed 2013-11-16] [WebCite Cache ID 6LAGieOLV]
23. Brenner J, Smith A. 72% of online adults are social networking site users. Washington, DC: Pew Internet & American Life Project; 2013 Aug 15. URL: http://pewinternet.org/Reports/2013/social-networking-sites.aspx [accessed 2013-11-16] [WebCite Cache ID 6LAGn7Z5c]
24. Smith A, Brenner J. Twitter use 2012. Washington, DC: Pew Internet & American Life Project; 2012 May 31. URL: http://www.pewinternet.org/Reports/2012/Twitter-Use-2012.aspx [accessed 2013-11-16] [WebCite Cache ID 6LAGvKrKN]
25. 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]
26. Lasswell H. The structure and function of communication in society. In: Bryson L, editor. The Communication of Ideas. New York: Harper; 1948.
27. Kreuter MW, McClure SM. The role of culture in health communication. Annu Rev Public Health 2004;25:439-455. [doi: 10.1146/annurev.publhealth.25.101802.123000] [Medline: 15015929]
28. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003 Dec;27 Suppl 3:S227-S232. [Medline: 14672383]
29. McGuire WJ. Theoretical foundations of campaigns. In: Rice RE, Atkin CK, editors. Public Communication Campaigns, Second Edition. Newbury Park, CA: SAGE Publications; 1989:43-65.
30. 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]
31. Duggan M, Brenner J. The demographics of social media users — 2012. Washington, DC: Pew Internet & American Life Project; 2013 Feb 14. URL: http://www.pewinternet.org/Reports/2013/Social-media-users.aspx [accessed 2013-11-16] [WebCite Cache ID 6LAGzpr6G]
32. Harris JK, Mueller NL, Snider D. Social media adoption in local health departments nationwide. Am J Public Health 2013 Sep;103(9):1700-1707. [doi: 10.2105/AJPH.2012.301166] [Medline: 23865660]
33. Smith M, Milic-Frayling N, Shneiderman B. CodePlex. 2010. NodeXL: Network Overview, Discovery and Exploration for Excel URL: http://nodexl.codeplex.com/ [accessed 2014-01-13] [WebCite Cache ID 6MbS0vJde]
34. Wu SS, Hofman JM, Mason WA. Who says what to whom on Twitter. 2011 Presented at: Proceedings of the 20th International Conference on World Wide Web; March 28-April 1, 2011; New York.
35. Lariscy RW, Avery EJ, Sweetser KD. An examination of the role of online social media in journalists. Public Relations Review 2009;35(3):314-316.
36. Cohen L, Chavez V, Chehimi S. Using media advocacy to influence policy. In: Prevention is Primary: Strategies for Community Well-being. San Francisco, CA: Jossey-Bass; 2010.
37. Harris JK, Brownson RC, Bell RA, Maier R, Cohen E, Mueller N. Twitter connections among local health departments as potential pathways for dissemination. 2014 Presented at: 47th Hawaii International Conference on System Sciences; 2014; Big Island, HI.
38. Harris JK, Cohen EL, Brownson RC. Examining the network of local health departments and their information sources on Twitter. 2013 Presented at: NPHIC National Conference on Health Communication, Marketing, & Media; 2013; Atlanta, GA.
39. Harris JK, Snider D, Mueller N. Social media adoption and use in health departments nationwide: the state of the states. Frontiers in Public Health Services and Systems Research 2013;2(1):e5 [FREE Full text]
40. Harris JK, Mueller NL, Snider D, Haire-Joshu D. Local health department use of twitter to disseminate diabetes information, United States. Prev Chronic Dis 2013;10:E70 [FREE Full text] [doi: 10.5888/pcd10.120215] [Medline: 23639765]
41. Harris JK, Maier RC, Jolani N. What's in a username? Finding local health departments on Twitter. Frontiers in Public Health Services and Systems Research 2013;2(4):5.
42. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. Health Promot Pract 2013 Mar;14(2):157-162. [doi: 10.1177/1524839912469378] [Medline: 23271716]
43. Twitter. 2013 Jan 11. @ChicagosMayor URL: https://twitter.com/ChicagosMayor/status/289790516839718913/photo/1 [accessed 2014-01-07] [WebCite Cache ID 6MS2KAW6r]

Abbreviations

NACCHO: National Association of County and City Health Officials

Edited by G Eysenbach; submitted 23.09.13; peer-reviewed by HP Lee, D Carvalho, R Thackeray; comments to author 11.11.13; revised version received 23.11.13; accepted 23.12.13; published 26.02.14.

Please cite as:
Harris JK, Choucair B, Maier RC, Jolani N, Bernhardt JM
Are Public Health Organizations Tweeting to the Choir? Understanding Local Health Department Twitter Followership
J Med Internet Res 2014;16(2):e31
URL: http://www.jmir.org/2014/2/e31/
doi:10.2196/jmir.2972
PMID:24571914

Original Paper

In the Words of the Medical Tourist: An Analysis of Internet Narratives by Health Travelers to Turkey

Margaret E Ozan-Rafferty[1*], RN, MHA, MBA, DHA; James A Johnson[2*], MSc, MPA, PhD; Gulzar H Shah[3*], MS, MSc Stats, PhD; Attila Kursun[4*], MD

[1]Dignity Health, Nevada Market, St. Rose Dominican Hospitals, Henderson, NV, United States

[2]Central Michigan Univeristy, Herbert H and Grace A. Dow College of Health Professions, Mount Pleasant, MI, United States

[3]Georgia Southern University, Jiann-Ping Hsu College of Public Health, Statesboro, GA, United States

[4]JADA Healthcare Services, LLC, King of Prussia, PA, United States

[*]all authors contributed equally

Corresponding Author:
Margaret E Ozan-Rafferty, RN, MHA, MBA, DHA
Dignity Health, Nevada Market
St. Rose Dominican Hospitals
832 Harbor Beach Court
Henderson, NV, 89002
United States
Phone: 1 3126595453
Fax: 1 7025227268
Email: meorafferty@gmail.com

Abstract

Background: Patients regularly travel to the West for advanced medical care, but now the trend is also shifting in the opposite direction. Many people from Western countries now seek care outside of their country. This phenomenon has been labeled medical tourism or health travel. Information regarding health travelers' actual outcomes, experiences, and perceptions is lacking or insufficient. However, advanced Internet technology and apps provide information on medical tourism and are a vehicle for patients to share their experiences. Turkey has a large number of internationally accredited hospitals, is a top tourism destination, and is positioning itself to attract international patients.

Objective: The objective of this research was to identify the important individual characteristics of health travelers, outline the push and pull factors for seeking health care in Turkey, identify satisfaction with the outcomes and the results of these individuals' treatments, and note positive and negative factors influencing their perceptions and overall experiences about patients' health travel.

Methods: This research uses qualitative data from Internet narratives of medical tourists to Turkey. Ethical considerations of using Internet narratives were reviewed. Narratives for analysis were obtained by using the Google search engine and using multiple search terms to obtain publicly posted blogs and discussion board postings of health travelers via purposeful sampling. Narratives were included if they were written in English, described travel to Turkey for health care, and were publicly accessible. Exclusion criteria included narratives that were on medical tourism facilitator or provider promotional websites, not in English, and did not describe an experience of a medical tourist. Medical tourists' written words were analyzed in an iterative analytic process using narrative analysis theory principles. Three stages of coding (open, axial, and selective) were conducted to identify characteristics and themes using qualitative analysis software.

Results: The narrative posts of 36 individuals undergoing 47 procedures who traveled to Turkey for medical care between 2007 and 2012 were analyzed. The narratives came from 13 countries, not including the narratives for which patient origin could not be determined. Travelers were predominantly from Europe (16/36, 44%) and North America (10/36, 28%). Factors driving travelers away from their home country (push factors) were cost and lack of treatment options or insufficient insurance coverage in their home country. Leading factors attracting patients to destination (pull factors) were lower costs, physician's expertise and responsiveness, and familiarity or interest in Turkey. Health travelers to Turkey were generally satisfied with the outcomes of their procedures and care provided by their physicians, many noting intent to return. Communication challenges, food, transportation, and gaps in customer service emerged as key areas for improvement.

Conclusions: This analysis provides an understanding of the insights of medical tourists through the words of actual health travelers. This nonintrusive methodology provides candid insights of common themes of health travelers and may be applied to study other patient experiences. The findings of this research expands the body of knowledge in medical tourism and serves as a platform for further qualitative and quantitative research on health travelers' experiences.

(J Med Internet Res 2014;16(2):e43) doi:10.2196/jmir.2694

KEYWORDS

medical tourism; Turkey; travel blogging; personal narratives; qualitative research; patient satisfaction; delivery of health care; globalization; social media internationality

Introduction

Background

Historically, patients have received health care in the country of their residence, but more recently, medical-related travel to other countries has substantially transformed this traditional mode of health care [1]. Although travel to receive medical care in another country is bidirectional, many developing countries are promoting their health care services to attract foreign patients [1-4].This phenomenon has been accelerated by increasing globalization, improved bilateral trade [5,6], patients' unmet health care needs, and increasing cost of services within the United States [7,8]. Travel for health care today has many forms. Although cosmetic procedures are popular, health travelers also seek open-heart surgery, hip or knee replacement [9-11], wellness treatments [12], and procedures yet to be approved in their home country [13]. Many estimates indicate medical tourism, also known as health travel [14], has significant economic potential worldwide [10,15], and many countries are positioning themselves to be providers of care and service for international patients [11,16-20].

The theoretical explanation of health travel may benefit from theories explaining human migration, although the former is different from migration in that migration involves somewhat permanent geographic relocation. Consistent with the push-pull theory of migration [21], moving from one country to another is typically influenced by factors that push individuals away from their country of origin, driving people to leave home, and factors that pull or attract people to a new country [22]. In their scoping review of the health tourists' experiences, Crooks et al [23] noted that both push and pull factors contributed to patient decision making when considering traveling for care.

Purpose of the Study

Information on patient's experiences of their care is a foundational aspect of health care quality [24]. Although surveys and interviews provide insights into care provided, over the past several years patients have begun to share their experiences on the Internet in the form of blogs, discussion posts, and tweets resulting in large amounts of data which may augment formal survey approaches [25] To date, there are few studies in the medical tourism literature that evaluate health travelers' experiences with care in another country [23]. Focusing on the narrative experiences of health travelers also provides feedback into the emotions associated with seeking care in an unfamiliar environment [26]. Limited accurate national data on the number, type, and impact of individuals who travel for health care have led to potentially inflated claims of industry growth [27] and made it difficult to address policy and public health needs of health travelers [28]. The majority of the published studies have focused on scoping reviews of the existing literature and identify research gaps [4,5,11,29,30]. Lack of detailed information on the common experiences of health travelers may result in inaccurate assumptions and misleading conclusions regarding their experiences and outcomes.

After an evaluation of the literature on health travel, Turkey was selected for this study because it has a large number (n=47) of internationally accredited health care providers [31] and is the world's sixth top tourism destination by arrivals [32]. Turkey boasts a geographic location attractive to European, Middle Eastern, African, and Asian health travelers [33], and is implementing multiple initiatives to make health and wellness care in Turkey attractive to international patients [34-36]. Although reports of the volume of health travelers vary, between 2008 and 2011 the number more than doubled according to the Directorate General of Health Services Department of Health Tourism, making the quality of health care in Turkey increasingly important on a global scale [34,37].

This qualitative narrative study explores the online narratives from health travelers to Turkey and describes the themes of their experiences. Qualitative research studies can capture an in-depth understanding of an issue [38]. Narrative research "begins with the experiences as expressed in lived and told stories of individuals" [38]. Although Internet and online documentation are relatively new phenomena, narrative analysis of online content has been conducted in several disciplines, including health care and travel [39-43].

It is expected that findings from this research will assist in decision making for patients considering health travel in the future, assist countries with their marketing and positioning to health travelers, and serve as a resource for hospitals wanting to recruit and retain staff to serve a global patient base. In addition, this research may strengthen health administration education by providing insights into the phenomenon of health travel.

Research Questions

The overarching question answered in this research is "What can we learn about health travelers to Turkey through analysis of their online narratives?" Additional subquestions included:

1. What are the important characteristics of health travelers who write online narratives about their experiences in Turkey?

2. What are the leading factors associated with the health travelers' country of origin that pushed them to seek health care in another country?

3. What are the leading factors associated with the destination country, Turkey, which pulled these health travelers to seek health care?

4. What can be derived from the narratives regarding travelers' satisfaction with the outcome and the result of their treatment they received in Turkey?

5. What are some positive and negative factors influencing health travelers' perceptions and overall experiences about their health travel to Turkey?

Methods

Search Strategy

Publicly available narratives were considered for this research after the ethical considerations of using online narratives were reviewed [44-49]. After receiving approval from the Central Michigan University Institutional Review Board, the narratives for analysis were obtained by using the Google search engine. Multiple search terms were used to obtain as many publicly posted narratives of health travelers to Turkey via purposeful sampling [50] during October 2012. The search utilized a range of terms, including health travel, medical tourism, surgery, and wellness travel. Clinical procedures used in the search were determined based on reviewing sites that promote health travel as well as findings from literature review and terms noted in selected posts, such as "I had surgery in Turkey."

The search was further refined by using Google's advanced search process in which the search terms were limited to discussions and blogs. Each result from the first 10 pages of these searches was reviewed to meet the criteria of: (1) a narrative written in English because all researchers were fluent in this language, (2) a first-person narrative written by an individual who underwent treatment and/or their significant other/partner on the trip, (3) included a description of the type of procedure, (4) a narrative that included a personal actual experience of health travel to Turkey, and (5) a narrative that was publicly available and did not require a password or discussion board membership to read. The authors reviewed all narratives to ensure promotional narratives found on health travel facilitator, promoter, or health care provider sites were excluded to avoid potential bias [28]. All narratives selected noted the name of the procedure performed and mentioned that this procedure was performed in Turkey.

Search Results

The initial Internet searches using 229 search terms resulted in 294,125,697 entries including those in languages other than English. Figure 1 denotes the narrative search process findings. Using the inclusion and exclusion criteria and taking a flexible approach to defining a narrative, both stories of the entire experience as well as those that were shorter in length highlighting a specific part of the health travel experience were included [41]. Narratives consisted of some discussion board posts of a few paragraphs on a single day to blogs that consisted of posts spanning several months with multiple entries. Only posts and blogs related to the topic of the author's health travel were included in the research.

After extensive searching, an initial 40 narratives were found that met the criteria. After a secondary review by the authors, 4 narratives were eliminated because further review indicated they may have been written by expatriates living in Turkey. A total of 36 narratives were deemed usable for the research, consisting of 23 message board or forum postings and 13 blogs.

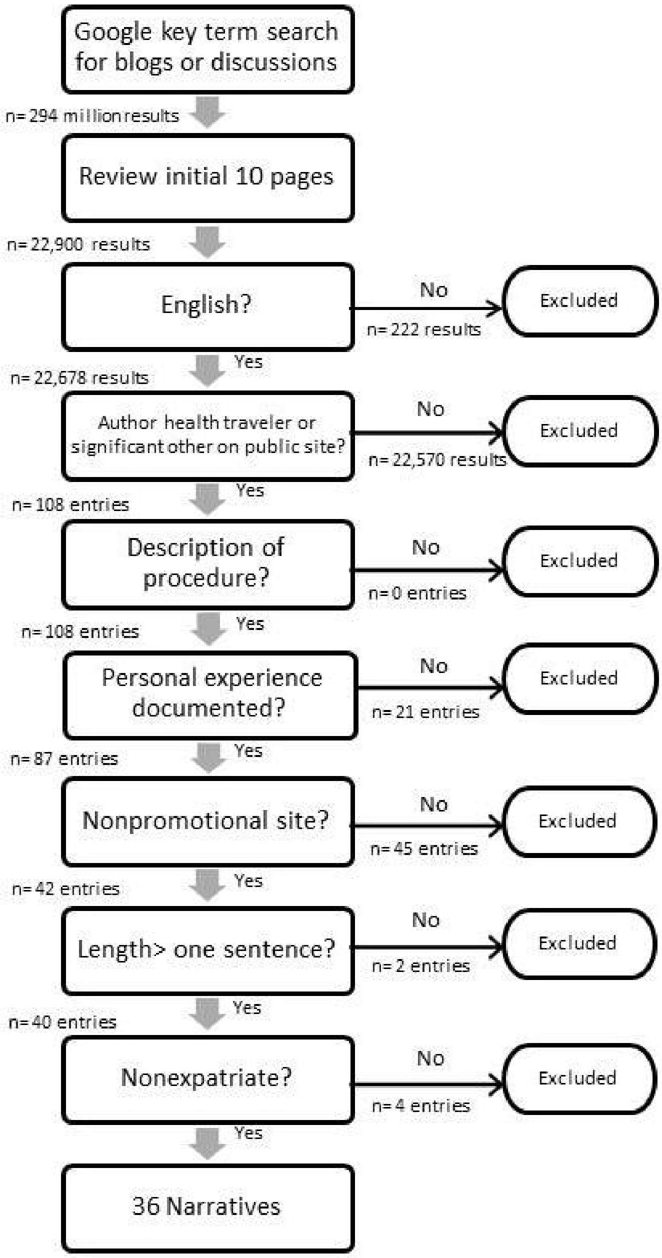
Coding Process

Posters' written words were analyzed in an iterative analytic process by using narrative analysis theory principles to create primary (parent) and secondary (child) themes. The unit of reference was a sentence [51,52]. Initial open coding was conducted on all pages of the 36 final posts to identify common themes and create initial codes for the data [50]. Each narrative was carefully reviewed to identify themes, with a focus on the initial research question and subquestions. In addition, the content of each sentence was directly examined to identify positive and negative opinions [52]. A total of 41 initial themes were identified and memos were written to keep track of ideas and observations. Two researchers, the coauthors of this paper, individually reviewed 5 cases and compared their agreements/disagreements on the major themes to assure level of intercoder reliability. Percent agreement on the 41 themes was 95.

Axial coding, or a second review of the data, was then completed and the 36 individual narratives were entered into NVivo version 10 [53] where the initial codes and key concepts identified in the open coding were distilled and clustering of the initial themes occurred [50]. Relationships with the various initial themes were created and characteristics of each case, where available, were noted.

A final round of selective coding [50] was completed by scanning all the data and codes, reviewing all themes and creating major themes (parent) and subthemes (child), and reorganizing the themes. During this review, linkages between some themes and the elimination of other initial themes by consolidation of themes were completed. Narratives from the posters representing each major theme were selected and personal details were removed from the quotes.

Figure 1. Steps in the narrative search process and findings.



Results

Health Travelers' Characteristics

The subjects (ie, posters used in the analysis) were identified by country of origin as shown in Table 1.

As shown in Table 2, the posters traveled to Turkey for a myriad of procedures. The 36 posters went for a total of 47 procedures. Dental work was the most frequent secondary procedure mentioned by the posters.

Demographic information, such as gender and age, is provided in Table 3 as well as some information on the facility or locale of the treatment. Of those who noted the place of their treatment, Anadolu Medical Center was mentioned most frequently. Most posters (17/36, 47%) noted that they had their procedure completed in Istanbul, 7 (7/36, 19%) went to Izmir, and 2 (2/36, 6%) traveled to Ankara. In all, 9 posters (9/36, 25%) mentioned a medical tourism facilitator organization by name and 3 (3/36, 8%) posters noted that they used the same facilitator.

Table 1. Health travelers’ country of origin (N=36).

Country	Number of travelers
North America (n=10, 28%)	
United States	8
Canada	2
Europe (n=16, 44%)	
United Kingdom	9
Netherlands	2
Finland	1
Germany	1
Ireland	1
Macedonia	1
Romania	1
Middle East (n=2, 6%)	
Dubai	1
Israel	1
Kenya (n=1, 3%)	1
Korea (n=1, 3%)	1
Undetermined (n=6, 16%)	6

Table 2. Health travelers’ primary and additional procedures performed in Turkey.

Procedure type	Procedure, n (%)		
	Primary procedure[a] (n=36)	Additional procedure (n=11)	Total (n=47)
Hair transplant	12 (26)		12 (26)
Dental	2 (4)	3 (6)	5 (11)
Abdominoplasty	3 (6)	1 (2)	4 (9)
Breast enlargement	3 (6)	1 (2)	4 (9)
Lasik	4 (9)		4 (9)
In vitro fertilization	2 (4)		2 (4)
Liposuction		2 (4)	2 (4)
Nerve surgery	2 (4)		2 (4)
Rhinoplasty	2 (4)		2 (4)
Back surgery	1 (2)		1 (2)
Bariatric	1 (2)		1 (2)
Botox and fillers		1 (2)	1 (2)
Cancer treatment	1 (2)		1 (2)
Cyst removal		1 (2)	1 (2)
Face lift		1 (2)	1 (2)
Heart surgery	1 (2)		1 (2)
Labioplasty		1 (2)	1 (2)
Physical exam	1 (2)		1 (2)
Stem cell transplant	1 (2)		1 (2)

[a] Primary procedure is the procedure posters noted first in their narrative.

Table 3. Health travelers' date of initial narrative, demographics, location of treatment, date of treatment, facilitator facility, and accompanying person.

Author	First post	Gender	Age	City	Date of treatment	Facilitator	Facility	Accompanying person
#1	7/22/10	M	30s	Izmir	5/1/10	—	—	—
#2	2/1/09	—	—	Istanbul	—	Blue Med travel	—	—
#3	4/1/12	M	—	Istanbul	—	—	Este of Turkey	—
#4	6/3/12	F	—	Izmir	6/1/12	—	Kent	—
#5	8/12/12	F	60s	Istanbul	10/12/12	Comfort Zone	—	—
#6	2/16/09	F	—	—	—	—	—	Spouse
#7	2/26/11	F	30s	Istanbul	3/12/11	World Med assist	Anadolu Hospital	Partner
#8	2/14/07	F	—	Izmir	—	Revitalize in Turkey	—	—
#9	9/13/12	F	—	—	—	—	Dunya Eye	Other referenced
#10	3/14/10	M	—	Ankara	2/22/10	—	Dr. Keser Clinic	—
#11	4/1/12	—	—	—	—	—	Este of Turkey	—
#12	6/7/11	M	50s	Istanbul	5/25/11	—	Transmed	—
#13	4/9/09	M	50s	Istanbul	5/1/09	World Med assist	Anadolu Hospital	—
#14	2/26/12	F	30s	Istanbul	5/1/12	IVF vacation center	Anadolu Hospital	Spouse
#15	5/22/11	F	—	Altinkum	—	—	—	Spouse
#16	N/A	M	30s	—	—	—	Hairana	—
#17	1/12/12	F	—	Izmir	12/12/11	Revitalize in Turkey	—	—
#18	1/14/09	F	—	—	1/2/09	—	—	—
#19	6/14/12	M	20s	Istanbul	9/1/11	—	Transmed	—
#20	10/17/12	M	30s	Istanbul	1/1/12	—	New Age Clinic	—
#21	5/7/08	M	20s	—	5/8/08	—	—	—
#22	10/9/07	M	—	Istanbul	8/9/07	—	Transmed	—
#23	8/30/12	M	20s	Ankara	8/1/12	—	Dr. Keser Clinic	—
#24	12/22/09	F	—	Istanbul	—	—	—	—
#25	7/4/12	F	40s	—	10/3/12	—	—	—
#26	4/8/11	M	40s	Izmir	5/19/11	—	—	Other family member
#27	9/24/12	F	—	—	5/1/12	—	—	—
#28	5/25/12	M	30s	Istanbul	5/12/12	—	Acibadem Hospital	—
#29	2/27/09	M	40s	Istanbul	4/1/09	World Med assist	Anadolu Hospital	Spouse
#30	5/12/12	M	30s	Istanbul	—	—	Hedz International	—
#31	1/29/12	F	60s	Istanbul	1/25/12	—	—	—
#32	5/3/12	F	—	—	9/1/11	—	—	—
#33	9/10/12	F	30s	Istanbul	10/12/12	—	Memorial Hospital	Spouse
#34	4/12/11	—	—	Izmir	6/19/12	—	—	—
#35	4/28/12	F	40s	Izmir	4/10/12	—	—	—
#36	2/2/12	F	—	Istanbul	—	Blue Med	—	—

Decision Analysis

Push Factors

The subjects of this study noted a variety of reasons, also known as *push factors*, for their decision to seek care outside of their home country. As noted in Table 4, push factors included lack of available treatment in their home country and financial reasons based on price or lack of insurance.

Table 4. Push and pull factors for patients who traveled to Turkey for medical procedures.

Factors	Discussion
Push factors	
Lack of treatment options in country of origin (3/36, 8%)	Several posters sought out treatment that was not available in their home country. These procedures included Cyber Knife treatment, stem cell transplants, and pudendal nerve surgery. These posters wrote about their disappointment with their inability to procure care in their own country.
Cost (22/36, 61%)	Posters note the price of procedures in their home country as rationale for seeking care in another country. Although only a few noted actual cost comparisons that included figures, many shared that they investigated costs abroad and noted that not only was the price of the procedure less than at home, many organizations offered packages which were all-inclusive and provided transportation, meals, and accommodations.
Lack of or insufficient insurance coverage (5/36, 14%)	The 3 US posters who went to Turkey for noncosmetic procedures, all noted that their surgeries were not covered by their insurance and, in one case, the patient did not have insurance. Two of these individuals had procedures which they noted were “experimental” in the United States, the other had heart surgery, but would have had to pay out of pocket because he lacked medical insurance coverage. One of the posters who traveled from the Netherlands for eye surgery noted that her insurance would only cover part of the cost of the procedure.
Pull factors	
Comparative value (24/36, 67%)	Posters shared their approach to learning more about Turkey and their procedures and generally noted that their ability to compare overall value of care (as defined by Porter and Teisberg [54]) available through alternatives options assisted with their decision making. The research performed by the posters was often detailed and time consuming. Many posters shared the options they investigated during their research process. The ability to connect with others who traveled to Turkey for information and support was also a component of some posters pretravel research.
Physician expertise and responsiveness (21/36, 58%)	Many posters mentioned the qualifications of the physicians as a factor that attracted them to care in Turkey. Some noted the background of the physician. One poster noted that it was the physician qualifications above costs that attracted them to Turkey. Other posters noted the communication or responsiveness of the physicians directly that attracted them to health travel to Turkey.
Familiarity or interest in Turkey (5/36, 14%)	Some of the posters noted that they had lived in Turkey, worked in Turkey, or been on vacation or honeymooned in Turkey and others noted an interest in exploring Turkey. Two of the posters noted they were of Turkish descent.
Availability of a health travel facilitator (13/36, 36%)	Some of the posters reached out to health travel facilitators as part of their process. Posters who mentioned using a health travel facilitator, someone typically on the ground in Turkey who serves in a sales consultative role, noted that the facilitators were responsive to their initial and online requests for information and assisted in the process of arranging treatments, transferring medical records, travel, and accommodations. Many of the facilitators were mentioned by name by the posters and several mentioned how quickly the facilitators responded to requests for information.
Price (23/36, 64%)	Many of the posters noted that they were attracted to Turkey for health travel because of the perceived reasonable cost of travel, visa entry, and price of treatment. The all-inclusive packages mentioned by the posters often included transportation, mobile phone support, meals, and accommodations in addition to the cost of the medical treatment.

Pull Factors

Table 4 also outlines a number of factors that influenced the posters’ decisions to seek health care in Turkey. Many of the posters noted researching their options, primarily on the Internet, and then either initiated contact with a provider/physician in Turkey directly or utilized a health travel facilitator for assistance. Price and a familiarity with Turkey were also factors that attracted posters to Turkey.

Satisfaction With Outcomes and Results of Treatment

Most posters (27/36, 75%) noted that they were satisfied with the outcome of their medical care in Turkey. Two hair transplant posters wrote negative reviews on a discussion board. One poster noted she was awaiting improvement in her nerve after surgery and the blogger who had end-stage cancer treatment died several months after treatment in Turkey.

Some posters (6/36, 17%) mentioned the impact that the procedure had on their lives. One of the posters who had been unable to stand up straight before his surgery wrote about the experience of being able to look up after his 12-hour surgery in Izmir:

> *On Sunday 22 May...my doctor asked if I was up to waking up from my bed and take a walk. What? Was he serious? I agreed and with the help of the nurses,*

> *I woke up from my bed with all the tubes hanging out. Boy did it feel weird taking my first steps. Felt like my back was in a vice. Walked to the windows and after so many years I was able to look up. You can't believe the feeling that I started to cry, yes, at my age but hell yes I cried with joy and victory.*

Very specific clinical outcome data including medical reports were noted as methods posters used to share the outcomes of their procedures. Several posters (7/36, 19%) offered to avail themselves via their personal email to anyone who needed additional information on their treatment and outcomes: "It has been 10 months since my last surgery and I have already had satisfying results that I would like to share with my fellow men who have been suffering from the same problem and hopefully I can be helpful as much as I can."

Several posters (4/36 11%) who were satisfied with their outcomes also mentioned an intent to return for additional procedures: "We will be going back there later this year as my husband still needs some work carried out to finish off his implants." Some of the posters who were less than satisfied with the outcome of their treatment also noted a desire to return in their posts as well. One couple who blogged about their failed in vitro fertilization attempts remained positive about their experiences in Turkey despite the negative outcome. Others were very direct in sharing their dissatisfaction and one even labeled their post "My disastrous journey with [facility]:"

> *This surgery with [facility] has made my life a living hell. I am not depressed but emotionally devastated.*

One poster who wrote about being satisfied with her results did mention an infection after her procedure, but she noted that she blamed herself for not strictly following her postoperative instructions. The wife of a poster made the final entry on his blog announcing his death several months after end-stage cancer treatment.

Perceptions and Overall Experiences

Positive Factors

Overview

The narratives included feedback on physicians, the facilitators, and the staff who cared for them or their loved one. Posters also wrote about the value they perceived, their perceptions of Turkey and the facilities where they received their care, and some gave details about their follow-up care.

Impression of Turkey

Most posters (10/11, 91%) who wrote about Turkey used favorable terms in their overall impressions. A couple posters compared Istanbul to other cities or their own country. Posters also noted their impressions of the people they met while in Turkey: "Turkish people are really friendly and were always willing to help out."

Approximately half of the posters (6/11, 55%) who wrote about Turkey mentioned taking or making the time to see the sights. For some of the posters, it was before their procedure: "...woke up early and went for a tour of the old city, [Topkapi] Museum, blue mosque, and basically walked about." For others, their sightseeing occurred after their procedures: "After the medical dressing and cleaning, I wore my hat and went for Istanbul city tour."

Perceptions of Physicians

Overwhelmingly, the authors mentioned their physicians in their postings (21/36, 58%) and of those who did, most of this feedback was positive (19/21, 91%). Most postings included the physicians by name (19/36, 53%) and many (14/36, 39%) mentioned the expertise of the physician:

> *Dr [name] was very friendly while informing me about the operation.*
>
> *I deeply felt his self-confidence during the consultation.*
>
> *He seems to do a lot of conferences, which is a good thing, I think, and has contributed to at least 2 of the papers being presented at this conference. He's keeping totally up to date on the progress in stem cell research, actively participates in cutting-edge research, and quite obviously uses his knowledge to treat his patients.*

Many posters (13/36, 36%) were also appreciative of the care and professionalism provided to them by their physicians.

Follow-Up

Several (5/36, 14%) posters mentioned the follow-up from their providers in Turkey:

> *I can reach the doctor with my questions by phone. I have also emailed them and got a reply within the same day.*

The posters noted that they received follow-up via email and one poster mentioned their physician's availability via Skype:

> *We have stayed in contact with Dr [name] at [facility] in [town] Turkey—email and Skype are wonderful ways to maintain communications, and it has been a real treat to actually see him and his office assistant and [name], our international rep from the medical center.*

Impression of Facilities

Many (7/36, 19%) posters noted the cleanliness of the facilities and some (3/36, 8%) commented on the surroundings and up-to-date equipment:

> *The hospital is bright and modern and clean.*
>
> *The private hospital where I got the operation was very pleasing with its full nursing facilities and advanced medical care units.*

Value

Many of the posters (23/36, 64%) noted that they felt that they "got their money worth" and were pleased with the value of the investment they made in their procedure(s): "Anyone needing hair transplants should think of Istanbul as the place; people and service are fine and at €1 a hair, it's incredible value." One couple seeking fertility treatment returned for a second round after the first in vitro fertilization in Turkey was unsuccessful.

Facilitators

Health travel facilitators were often mentioned by name (7/36, 19%) and evaluated by some for their assistance and responsiveness during the posters time in Turkey. These individuals greeted the posters at the airport and provided support during the stay and also facilitated site seeing etc: “When I arrived at the airport in Turkey, Dr [name]’s English translator and the German translator were waiting for me.”

Some (4/36, 11%) posters noted that the facilitators provided a direct contact number and 2 posters noted that they also received a phone subscriber identity module (SIM) card or telephone to aid in communication. “They took care of the check-in and also gave me a prepaid SIM card in case I had to contact them for any reason.”

Negative Factors

Overview

Two posters (2/12, 17%) were not satisfied with their hair growth after transplants. Several posters (7/36, 19%) noted some areas of dissatisfaction with their experience in Turkey and identified opportunities for improvements. Overwhelmingly, these negative posts focused on issues with communication and the lack of ability to speak to their caregivers in English. Food, transportation, and responsiveness were also mentioned as opportunities for improvement and several gaps in service were noted by posters who documented these details of their experiences: “I contacted [the provider] but my query was ignored—I did not get any feedback.”

Communication

Although many posters (8/36, 22%) described lengthy communications with their physician or the physician’s ability to speak to them in English, many posters (7/36, 19%) noted the challenges of communicating with the staff and support team members:

> *Language was a problem as we are from an English-speaking country, but did get along with a lot of hand language.*
>
> *Whilst Dr [name] speaks English, most of the nursing staff did not, so I got by with sign language and a few words.*

Several posters (4/36, 11%) noted an inability to communicate with the staff, such as nurses, and struggled to make their needs known when dealing with clinical needs: “Today’s challenge: How does one communicate, ‘I’m constipated’ with folks who don’t understand English?”

Challenges with others, such as taxi drivers, were also noted:

> *As we climbed in the taxi to take us home, I told the driver the name of our hotel. He said “okay,” and we were off. About halfway there he took a wrong turn (I don’t know many streets, in Istanbul, but I knew this one). A minute later, he pulled up to the wrong hotel. We said, “uhhh, this isn’t it.” He took us to one with a similar name. After that he gave us a pretty big lecture in Turkish about always getting a hotel card when we leave.*

Food

Many of the posters who wrote about food (3/5, 60%) mentioned their challenges with having the types of food they were used to having at home: “Fruit has been not available since I got here, so bought a large bunch of grapes on my way back to the apartment.”

Those who wrote about the food served to them during their hospitalizations shared their feedback on items they were served during their stay: “Afterwards, I did manage to sleep until around 6:30 am when they brought me my breakfast—a light salad with some goat’s cheese and olives (yuck) and a cup of black tea.”

One poster who raved about her care and outcome noted the only issues she had with her stay:

> *Only complaint is the food SUCKED. I was only given soup for 2 days and consistency was like baby food. I had the lady helping me get some instant soup at the supermarket, much better!!*

Transportation

Many of the posters who wrote about their experiences noted their perception of the local traffic (3/10, 33%) “The drive to the hospital was a little intense, in Turkey they are not good with following the traffic laws, so everyone seems to do what they want, a little scary.” Traffic was a barrier to one poster who was considering seeing more of the sights in Istanbul: “Traffic in Turkey is like nothing we have ever experienced, instead of trying to find a taxi and having a stressful drive to the sights, and not being sure what is and isn’t open on Sunday, staying in seemed to be the logical choice.” Two of the posters from the United States noted the multiple flights they took to get to Turkey and the challenges of traveling this distance.

Impressions of the People

Although most comments about the staff and general population were positive (5/7, 71%), some (2/7, 29%) employee behaviors observed by the posters were less than positive:

> *Monday morning, I was looking out the window as folks were coming in to work. Lots of cigarette smokers here and it’s not uncommon to see them outside the entrances to the hospital. It didn’t take me long, though, to figure out this guy was smoking more than a cigarette. One toke over the line, sweet Jesus...*

Other negative touch points with individuals involved in the treatment were also noted:

> *My dealings were mostly with another rep, [name], who is obviously very busy juggling many patient/doctor schedules at any one time. However, maybe it was just me, but I felt my interactions with her were sometimes a little “edgy” and I was not always happy when what I considered were legitimate doubts or concerns were met with what appeared to be impatience and touchiness. Maybe it’s a cultural thing, but it wasn’t what I am used to by way of customer care, and I didn’t find it very reassuring at all.*

Responsiveness

For the most part, posters who mentioned communicating with the facilities or providers were satisfied with their responses (6/9, 67%). However, one person who was dissatisfied with the outcome of his hair transplant noted that he repeatedly tried to get a response from the provider, but needed several attempts before he got a satisfactory response. Another poster noted that he attempted to reach an organization to arrange a health physical but did not receive a reply: “I contacted [facility] first but they neglected to return my call.”

Unexpected Service Gaps

Several posters (4/36, 11%) noted omissions in service or occurrences of incidents they did not expect. One patient arrived at the clinic expecting to be treated by a specific physician, only to be assigned an alternative physician:

> *When I arrived to the clinic on the day of my surgery, I learned that Dr [Name] cannot perform FUE [follicular unit extraction] surgeries and even more she was out of the country...and instead it will be done by a different surgeon, Dr [name].*

One of the posters shared that she incurred unanticipated costs associated with her treatment:

> *We dealt with some frustrations regarding procuring medications once she was released as an in-patient. Unlike US hospitals, which often have pharmacies where patients can buy their drugs, in Turkey, it’s illegal for hospital pharmacies to “sell” drugs to those who are not inpatients. So we ended up going to the “outside” pharmacy for her meds for this week and found ourselves unexpectedly paying a lot of money for a very expensive drug that we thought was going to be covered in the pretransplant costs that have already been paid for.*

One poster and her partner’s summary of their final hours in Turkey revealed her challenges as she tried to get to her flight:

> *We waited at the entrance to the airport for about a half hour while our driver went to get a wheelchair for [name]. He came back empty handed, which meant we needed to load all the luggage on one cart, and push it through security before we could get to a check-in counter.*

Three of the posters mentioned that they were frustrated with having only a single English-speaking television channel and one mentioned being unable to play her US DVDs in the DVD player. A poster who stayed for an extended hospitalization noted an inability to do her laundry on-site and having to send items out for cleaning.

Other Negative Factors

One of the posters went to into great detail about her frustrations with getting mail sent to her in Turkey. She notes the post office demanding payment to secure a package:

> *The Turkish mail system doesn’t have a great reputation, so our own postal system loses control once it leaves the US.*

In addition, one poster noted that when out and about in the outskirts of Istanbul, the poster noted their thoughts on air pollution and suggested to readers to avoid going outside:

> *Not great air quality here because of the factories so close by so going out walking too much isn’t a great idea.*

The partner of the patient who stayed over 2 months in Turkey noted that after a few days her blog was censored and she had to email her posts to a friend to post from the United States:

> *This blog has been censored by the Turkish government so, being situated in Turkey, I can’t access it. So I’m going to try an end run by writing and then forwarding to someone in the US who can upload postings.*

Discussion

Principal Findings

This analysis of online narratives provides significant insights into experiences of health travelers to Turkey. This study provides information from 36 individuals who posted their characteristics, the factors that drove them to leave their home counties for care, and what attracted them to seek care in Turkey. Details from the posters on the outcomes of their procedures and their satisfaction with their experiences in Turkey provide an understanding of both positive and negative factors influencing their perception of health travel to Turkey. Overall, these insights also provide individuals exploring options for health care abroad with information about others’ experiences as health travelers and may aid in the decision process of those seeking care in another country. Marketers of health travel may benefit from our analysis as they develop strategies to address the influence that the insights and opinions of current health travelers have on individuals considering future travel for care. As consumers of health care seek information on treatment options and outcomes, online outcome data and reviews may also influence their decisions of care choices.

This purposeful sample of 36 individual posters of narratives mirrors the size of samples analyzed by other researchers [42,49]. After careful reading and rereading of the narratives, we felt confident that the sample contained posts by actual health travelers who went abroad to Turkey for care and those entries by medical tourism promotional sites or health care providers were eliminated. Many of these narratives provided rich insights into the rationale for traveling to Turkey for care as well as detailed feedback regarding their thoughts and perceptions before, during, and after treatment. As noted by Seale et al [55], the narratives in the study provided a great amount of detail about the experience, including technical and intimate details and observations.

Posters of all ages traveled from many countries across the globe seeking care in Turkey for procedures that were unavailable or unaffordable in their home country, or not covered by insurance. This finding aligned with other studies on the drivers of patients looking for health care abroad [23,56,57]. Although wait time for procedures is noted as a push factor for some individuals seeking care abroad [4,10,17,58], none of the posters wrote that they were waiting for care in their home country.

Posters who traveled for care unavailable in their home countries had common frustrations with their current health care options, or lack thereof, and highlighted their desperation and the challenges faced by individuals seeking a solution for their illness or pain. Posters were often involved in lengthy research processes used to explore options for care with the Internet serving as the primary source of information and knowledge sharing.

Most of the authors of the narratives in this study were pulled to travel to Turkey for care that was less expensive than the treatment in their home country. Posters who traveled to Turkey for more affordable health care options also noted performing research before making the decision to travel outside their home country for care. Posters wrote about finding the best value through research on the Internet via reading other's narratives and communicating with providers or health travel facilitators. As noted by Crooks et al [23], positive stories of success via online postings or word-of-mouth may serve as a motivator for individuals considering health travel. These posters' research included information about the facility and investigating the physician's qualifications and previous outcomes. Many of these posters mentioned the value of the treatment they received.

Over one-fifth of the posters (8/36, 22%) mentioned traveling with a spouse or family member and in some cases the spouse or family member also contributed to the narrative. These accompanying persons typically posted when the primary poster was undergoing a procedure and, in one bloggers case, after he died. Although accompanying care persons may require additional attention, resources, and have exposure to the stress of travel and caregiving, they are often key stakeholders in a health traveler's support team [59]. Having these individuals included in the postings gave insights into the perspectives of the person who was the patient, but also their support person's view on the experience.

These 36 narratives were written primarily from the perspective of individuals who traveled for health rather than by those interested in combining a vacation with medical care. Although travel and the ability to see the sights in Turkey was noted by some of the posters, many did not mention taking a vacation or traveling to tourist location(s) and focused their posts on the clinical and medical procedures rather than an overseas holiday.

Accreditation organizations, such as Joint Commission International, were only mentioned by 3 of the posters in parts of their narratives related to their decision-making process to choose Turkey for care. Although frequently promoted on health travel websites [28], as the number of accredited organizations focused on health travel increase, international accreditation may not be a long-term source of competitive advantage for providers seeking to care for health travelers. As accreditation becomes an expectation, creation of high levels of brand trust via marketing and loyalty through positive word-of-mouth are likely to become the key drivers of a health travel provider's success [60,61]. In addition, although several posters mentioned their long flights to Turkey, proximity was only mentioned by one poster as a pull factor.

Most posters provided details on their procedures, outcomes, the care they received, the facilities, and their impression of Turkey, and these aspects were often imbedded into their narratives. The physicians and medical tourism facilitators were most often mentioned by the health travelers and frequently by name. Several posters noted that the accessibility of their physicians by email or phone after they returned home was important. Posters who used a health travel facilitator commented on the responsiveness and accessibility of the facilitator both before and during their visits to Turkey. These findings may be useful for future health travelers.

Few completely negative narratives were found during the extensive search for blogs and discussion board postings (3/40, 8%). Two posters noted their dissatisfaction with the outcome of their hair transplants and used their narratives as a vehicle to communicate their disappointment and frustrations.

In contrast to the physicians, the nursing staff were rarely mentioned by name and many posters shared their frustrations with being unable to communicate with the individuals directly responsible for their day-to-day care. The inability to make basic needs known to the staff, such as requests for water and help with elimination, has both patient safety and satisfaction implications that need to be addressed.

Opportunities also exist to improve various other touch points in the care of international patients in Turkey. These include improving the perception of the food served, transportation providers (especially taxis), and addressing unanticipated service gaps, such as an inability to find English-language television programs and lack of laundry services. Food was a dissatisfying element for several posters who mentioned their meals in their narratives. There is a need to address the cultural food preferences of health travelers and provide a more customized approach to menus and meals especially in the hospital setting.

Communication challenges with service personnel in Turkey, such as taxi drivers or merchants, were also noted. In one case, the poster articulated that she deferred sightseeing during her stay because of concerns about using a taxi in Istanbul. Posters also mentioned several incidents in which staff or other service providers demonstrated unprofessional behaviors. In addition, one poster noted the blocking of her blog during her stay in Turkey which resulted in her creating a work-around to be able to continue to communicate her experiences during hospitalization. The negative perceptions of government entities, such as the postal service, Internet censorship, or air quality in the city of care, although possibly isolated incidents, are additional opportunities for improvement.

Much like the growing body of patients who are sharing their experiences via Facebook and other sites, [62] these health travelers to Turkey have embraced the use of blogs, forum postings, and discussion boards to highlight their experiences, often in great detail. This research has demonstrated that analysis of online narratives provides a comprehensive review and insights into the experiences of health travelers to Turkey.

Limitations

Although a number of processes were put in place to ensure the authenticity of the narratives, it was impossible to guarantee that all the narratives in this study were trustworthy. The narratives may have included fictitious data and there was no way for the researchers to validate most of the self-reported information [48]. In addition, although extensive focus was taken to ensure the narratives were genuine and nonpromotional in nature, it was impossible for the researchers to ensure that the blogs and forum posts reviewed were not created at the urging of a provider or heath travel facilitator. However, many posters shared both positive and less-than-positive experiences in their narratives and often provided both criticism and positive feedback in the same post. In addition, the posters did not always note their gender, country of origin, or other demographic information. However, anonymity may have also encouraged people to share more online and discuss sensitive topics and issues [51] making the 36 narratives a robust source of information. The narratives lacked examples of any orthopedic cases, such as individuals seeking knee or hip replacements, or any individuals who wrote that they were waiting for care in their home country.

Although the number of people with access to the Internet and the ability to post has increased worldwide [63], the sample did not include individuals who did not chose to create online narratives about their experiences [64] or those who lacked access to a computer or smartphone; thus, they may not be representative of the general population [55]. In addition, using the Internet as a source excludes the ability to obtain the benefits of the spoken word, such as inflection and other nuances, as well as visual cues from personal interviews [55]. The study also excludes other sources of patient feedback, such as formal surveys and word-of mouth feedback [60]

Our research was limited to the English language only. The majority of foreigners seeking and receiving health care in Turkey are from countries where English is not the main language [35]. Therefore, our results may not represent the complete spectrum of health travel experiences in Turkey. More specifically, culturally biased experiences, such as language, overall communication, and food, may not apply to all health travelers visiting Turkey.

Conclusions

This research is believed to be the first of its kind in its approach to analyzing the online narratives of health travelers. This analysis provides an understanding of the insights of health travelers through the words of actual health travelers. The findings of this research expands the body of knowledge in medical tourism as well as serves as a platform for further qualitative and quantitative research on health travelers' experiences.

The nonintrusive approach of the methodology used has provided candid insights into the experiences of health travelers. As transparency in health care increases and patient satisfaction data and feedback becomes more publicly available, this methodology could be applied to study other patient experiences in various health care settings.

As an increasingly important destination for health travelers, positive attributes about Turkey include the expertise and responsiveness of physicians, clinical facilities, overall satisfaction with the outcome of the procedure, and overall impressions of Turkey. Negative attributes include challenges with communication with the nonphysician staff including nurses and assistants, food, traffic and several service gaps.

Providers of international patient care may use patient experience research in positioning their services and in the development of patient care protocols for their health travelers. Additionally, hospitals may integrate patient experience research discoveries into their employee recruitment and training programs. In the future, health administration programs may reference expanded outcomes of this research as they evaluate their curricula and may decide to include additional classes on health travelers and their experiences. In the long term, this research may serve as a platform for the development of an international forum of health traveler experiences

Authors' Contributions

This research was conducted as part of the first authors' requirements for completing a Doctorate of Health Administration at Central Michigan University. The second author served as dissertation chair and the third and fourth authors served as dissertation committee members.

Conflicts of Interest

None declared.

References

1. Ramírez de Arellano AB. Medical tourism in the Caribbean. Signs (Chic) 2011;36(2):289-296. [Medline: 21114073]
2. Carrera PM, Bridges JF. Globalization and healthcare: understanding health and medical tourism. Expert Rev Pharmacoecon Outcomes Res 2006 Aug;6(4):447-454. [doi: 10.1586/14737167.6.4.447] [Medline: 20528514]
3. Forgione DA, Smith PC. Medical tourism and its impact on the US health care system. J Health Care Finance 2007;34(1):27-35. [Medline: 18972983]
4. Hopkins L, Labonté R, Runnels V, Packer C. Medical tourism today: what is the state of existing knowledge? J Public Health Policy 2010 Jul;31(2):185-198. [doi: 10.1057/jphp.2010.10] [Medline: 20535101]
5. Smith R, Martínez Álvarez M, Chanda R. Medical tourism: a review of the literature and analysis of a role for bi-lateral trade. Health Policy 2011 Dec;103(2-3):276-282. [doi: 10.1016/j.healthpol.2011.06.009] [Medline: 21767892]

6. Turner LG. Quality in health care and globalization of health services: accreditation and regulatory oversight of medical tourism companies. Int J Qual Health Care 2011 Feb;23(1):1-7 [FREE Full text] [doi: 10.1093/intqhc/mzq078] [Medline: 21148210]
7. Runnels V, Carrera PM. Why do patients engage in medical tourism? Maturitas 2012 Dec;73(4):300-304. [doi: 10.1016/j.maturitas.2012.08.011] [Medline: 23007007]
8. Crooks VA, Cameron K, Chouinard V, Johnston R, Snyder J, Casey V. Use of medical tourism for hip and knee surgery in osteoarthritis: a qualitative examination of distinctive attitudinal characteristics among Canadian patients. BMC Health Serv Res 2012;12:417 [FREE Full text] [doi: 10.1186/1472-6963-12-417] [Medline: 23170924]
9. Alleman BW, Luger T, Reisinger HS, Martin R, Horowitz MD, Cram P. Medical tourism services available to residents of the United States. J Gen Intern Med 2011 May;26(5):492-497 [FREE Full text] [doi: 10.1007/s11606-010-1582-8] [Medline: 21161425]
10. Bookman MZ, Bookman KR. Medical Tourism in Developing Countries. New York: Palgrave Macmillan; 2007.
11. Horowitz MD, Rosensweig JA. Medical tourism--health care in the global economy. Physician Exec 2007;33(6):24-6, 28. [Medline: 18092615]
12. Boulay R, Hritz NM, Ashton C. An exploratory study of wellness travel: differences between US and non-US travelers. Journal of Tourism Insights 2013 Jul 23;4(1):1-18 [FREE Full text] [doi: 10.9707/2328-0824.1036]
13. 60 Minutes. 2005 Apr 35. Vacation, Adventure and Surgery: Elective Surgeries By World-Class Doctors At Third-World Prices URL: http://www.cbsnews.com/news/vacation-adventure-and-surgery/ [accessed 2014-01-26] [WebCite Cache ID 6Mve6nlSv]
14. Hofer S, Honegger F, Hubeli J. Health tourism: definition focused on the Swiss market and conceptualisation of health(i)ness. J Health Organ Manag 2012;26(1):60-80. [Medline: 22524099]
15. Deloitte Center for Health Solutions. Medical Tourism Update and Implications URL: http://www.deloitte.com/view/en_US/us/industries/US-federal-government/center-for-health-solutions/health-care-consumerism/55d9f278c9184210VgnVCM200000bb42f00aRCRD.htm [accessed 2013-04-23] [WebCite Cache ID 6G7WVRMtG]
16. Bozkurt N. Hurriyet Daily News. 2011 Nov 17. Turkey attracting more patients, minister says URL: http://www.hurriyetdailynews.com/ [accessed 2013-04-24] [WebCite Cache ID 6G7Wo85nc]
17. The Economist Intelligence Unit. 2011. Travelling for health the potential for medical tourism URL: https://www.eiu.com/public/topical_report.aspx?campaignid=MedTourism_Sept11 [accessed 2014-01-26] [WebCite Cache ID 6Mveq7x17]
18. Gupta V, Das P. Medical tourism in India. Clin Lab Med 2012 Jun;32(2):321-325. [doi: 10.1016/j.cll.2012.04.007] [Medline: 22727009]
19. Medical Tourism Magazine.: M; 2011 Jun 11. Medical tourism providing needed band-aid to Turkish debt URL: http://medicaltourismmag.com/medical-tourism-providing-nedded-band-aid-to-turkish-debt/ [accessed 2014-01-27] [WebCite Cache ID 6MvovhTjr]
20. Sengupta A. Medical tourism: reverse subsidy for the elite. Signs (Chic) 2011;36(2):312-319. [Medline: 21114076]
21. Lee ES. A theory of migration. Demography 1966;3(1):47-57 [FREE Full text] [doi: 10.2307/2060063]
22. Globalization 101. 2012. Why does migration happen URL: http://www.globalization101.org/why-does-migration-happen/ [accessed 2012-05-25] [WebCite Cache ID queryurlhttp3A2F2Fwwwglobalization101org38dat09e20130424]
23. Crooks VA, Kingsbury P, Snyder J, Johnston R. What is known about the patient's experience of medical tourism? A scoping review. BMC Health Serv Res 2010;10:266 [FREE Full text] [doi: 10.1186/1472-6963-10-266] [Medline: 20825667]
24. Institute of Medicine of the National Academies. 2001 Mar 01. Crossing the quality chasm: a new health systems for the 21st century URL: http://www.iom.edu/Reports/2001/Crossing-the-Quality-Chasm-A-New-Health-System-for-the-21st-Century.aspx [accessed 2013-12-18] [WebCite Cache ID 6Lxa8Ul9m]
25. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
26. Kingsbury P, Crooks VA, Snyder J, Johnston R, Adams K. Narratives of emotion and anxiety in medical tourism: on State of the Heart and Larry's Kidney. Social & Cultural Geography 2012 Jun;13(4):361-378. [doi: 10.1080/14649365.2012.683807]
27. Connell J. Contemporary medical tourism: conceptualisation, culture and commodification. Tourism Management 2013;34:1-13 [FREE Full text]
28. Lee H, Wright KB, O'Connor M, Wombacher K. Framing medical tourism: an analysis of persuasive appeals, risks and benefits, and new media features of medical tourism broker websites. Health Commun 2013 Oct 18:1. [doi: 10.1080/10410236.2013.794412] [Medline: 24138286]
29. Johnston R, Crooks VA, Snyder J, Kingsbury P. What is known about the effects of medical tourism in destination and departure countries? A scoping review. Int J Equity Health 2010;9(1):24 [FREE Full text] [doi: 10.1186/1475-9276-9-24] [Medline: 21047433]
30. Lunt NT, Mannion R, Exworthy M. A framework for exploring the policy implications of UK medical tourism and international patient flows. Social Policy & Administration 2012 Jan 18;47(1):1-25. [doi: 10.1111/j.1467-9515.2011.00833.x]

31. Joint Commission International. 2012. Accredited organizations: hospitals URL: http://www.jointcommissioninternational.org [accessed 2012-05-25] [WebCite Cache ID 6Mvjz3NYm]
32. UNWTO Tourism Highlights, 2013 Edition. Madrid, Spain: UNWTO; 2013. URL: http://dtxtq4w60xqpw.cloudfront.net/sites/all/files/pdf/unwto_highlights13_en_hr.pdf [accessed 2014-01-26] [WebCite Cache ID 6MvkEHgZG]
33. Altın U, Bektaş G, Antep Z, İrban A. The international patient's portfolio and marketing of Turkish health tourism. Procedia - Social and Behavioral Sciences 2012 Oct;58:1004-1007 [FREE Full text] [doi: 10.1016/j.sbspro.2012.09.1080]
34. Republic of Turkey, Ministry of Health, Directorate of Health Services. Evaluation Report on Medical Tourism in Turkey. 2012. URL: http://www.saglik.gov.tr/SaglikTurizmi/dosya/1-79232/h/eng-medikal-turizm.pdf [accessed 2014-01-26] [WebCite Cache ID 6MvksZCVR]
35. Republic of Turkey, Ministry of Health, Directorate of Health Services. Medical Tourism 2012 Activity Report. 2012. URL: http://www.saglik.gov.tr/SaglikTurizmi/dosya/1-79250/h/enfaaliyetraporu2012.pdf [accessed 2013-12-17] [WebCite Cache ID 6LxcZMyIO]
36. Erdoğan S, Yilmaz E, İşletmecilği T. Medical Tourism: An Assessment on Turkey. 2012 Presented at: Joint Conference of the 11th International Conference of the Asia Chapter & 2nd Conference of the MENA Chapter of the AHRD & 10th International Conference on Knowledge, Economy and Management; November 08-10, 2012; Istanbul, Turkey URL: http://www.medeniyet.edu.tr/content/userfiles/Medical_Tourism_An_Assessment_on_Turkey_pdf.pdf
37. Health Foundation of Turkey. 2010. Turkish and global perspective: Health tourism 2010 URL: http://www.saglik.org.tr/en/report/health-tourism-report-2010-1.html [accessed 2014-01-26] [WebCite Cache ID 6MvoAWCWT]
38. Creswell JW. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Thousand Oaks, CA: Sage Publications Inc; 2006.
39. Bosangit C, McCabe S, Hibbert S. What is told of travel blogs? Exploring travel blogs for consumer narrative analysis. In: Hopken W, Gretzel U, Law R, editors. Information and Communication Technologies in Tourism 2009: Proceedings of the International Conference in Amsterdam, The Netherlands, 2009. New York: Springer-Verlad; 2009.
40. Kelleher C, Helkkula A. Virtually speaking-customer to customer communication in blogs. J Applied Management and Entrepreneurship 2010 Jul 01;15(3):4-17.
41. O'Brien MR, Clark D. Unsolicited written narratives as a methodological genre in terminal illness: challenges and limitations. Qual Health Res 2012 Feb;22(2):274-284. [doi: 10.1177/1049732311420737] [Medline: 21876209]
42. Rachul C. What have I got to lose?: an analysis of stem cell therapy patients blogs. Health Law Rev 2011;20(1):5-12 [FREE Full text]
43. Tussyadiah IP, Fesenmaier DR. Marketing places through first‐person stories-an analysis of Pennsylvania Roadtripper blog. Journal of Travel & Tourism Marketing 2008 Dec;25(3-4):299-311. [doi: 10.1080/10548400802508358]
44. Clarke J, van Amerom G. A comparison of blogs by depressed men and women. Issues Ment Health Nursing 2008 Mar;29(3):243-264. [doi: 10.1080/01612840701869403] [Medline: 18340611]
45. Eastham LA. Research using blogs for data: public documents or private musings? Res Nurs Health 2011 Aug;34(4):353-361. [doi: 10.1002/nur.20443] [Medline: 21626517]
46. Heilferty CM. Ethical considerations in the study of online illness narratives: a qualitative review. J Adv Nurs 2011 May;67(5):945-953. [doi: 10.1111/j.1365-2648.2010.05563.x] [Medline: 21231955]
47. Hookway N. Entering the blogosphere: some strategies for using blogs in social research. Qualitative Research 2008 Feb 01;8(1):91-113. [doi: 10.1177/1468794107085298]
48. Jones M, Alony I. Blogs: the new source of data analysis. Issues in Informing Science and Information Technology 2008;5:433-446 [FREE Full text]
49. Pitts V. Illness and Internet empowerment: writing and reading breast cancer in cyberspace. Health (London) 2004 Jan;8(1):33-59. [doi: 10.1177/1363459304038794] [Medline: 15018717]
50. Neuman WL. Social Research Methods: Qualitative and Quantitative Approaches. Boston: Pearson/Allyn and Bacon; 2006.
51. Pfeil U, Zaphiris P. Applying qualitative content analysis to study online support communities. Univ Access Inf Soc 2009 Aug 1;9(1):1-16. [doi: 10.1007/s10209-009-0154-3]
52. O'Leary DE. Blog mining-review and extensions: “From each according to his opinion”. Decision Support Systems 2011 Nov;51(4):821-830. [doi: 10.1016/j.dss.2011.01.016]
53. QSR International. 2010. NVIVO v 10 URL: http://www.qsrinternational.com/ [WebCite Cache ID queryurlhttp3A2F2Fwwwqsrinternationalcom2F38date20130424]
54. Porter ME, Teisberg EO. Redefining Health Care, Creating Value-based Competition on Results. Boston: Harvard Business School Press; 2006.
55. Seale C, Charteris-Black J, MacFarlane A, McPherson A. Interviews and internet forums: a comparison of two sources of qualitative data. Qual Health Res 2010 May;20(5):595-606. [doi: 10.1177/1049732309354094] [Medline: 20008955]
56. Steklof CD. Medical tourism and the legal impediments to recovery in cases of medical malpractice. Wash U Global Stud L Rev 2010;9:721-742 [FREE Full text]
57. Veerasoontorn R, Beise-Zee R. International hospital outshopping: a staged model of push and pull factors. International Journal of Pharmaceutical and Healthcare Marketing 2010;4(3):247-264. [doi: 10.1108/17506121011076174]

58. KPMG International. Issues Monitor Healthcare: Medical Tourism Gaining Momentum. 2011 May. URL: http://www.kpmg.com/CH/en/Library/Articles-Publications/Documents/Sectors/pub-20120207-issues-monitor-healthcare-medical-tourism-en.pdf [accessed 2012-05-25] [WebCite Cache ID 20130424]
59. Casey V, Crooks VA, Snyder J, Turner L. You're dealing with an emotionally charged individual: an industry perspective on the challenges posed by medical tourists' informal caregiver-companions. Global Health 2013;9:31 [FREE Full text] [doi: 10.1186/1744-8603-9-31] [Medline: 23889860]
60. Yeoh E, Othman K, Ahmad H. Understanding medical tourists: word-of-mouth and viral marketing as potent marketing tools. Tourism Management 2013 Feb;34:196-201. [doi: 10.1016/j.tourman.2012.04.010]
61. Lertwannwit A, Gulid N. International tourists'; service quality perception and behavioral loyalty toward medical tourism in Bangkok metropolitan area. The Journal of Applied Business Research 2011 Dec;27(6):0892-7626.
62. Health Strategy Innovation Cell, The Change Foundation. Using Social Media to Improve Healthcare Quality: A Guide to Current Practice and Future Promise. Toronto: The Change Foundation; 2011. URL: http://www.changefoundation.ca/library/using-social-media-to-improve-healthcare-quality-part-1/ [WebCite Cache ID 6Jww3BxJC]
63. International Telecommunication Union. 2012. Data release: key statistical highlights URL: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2013-12-01] [WebCite Cache ID queryurlhttp3A2F2Fwwwituint38date20130424]
64. Marcus MA, Westra HA, Eastwood JD, Barnes KL, Mobilizing Minds Research Group. What are young adults saying about mental health? An analysis of Internet blogs. J Med Internet Res 2012 Jan;14(1):e17 [FREE Full text] [doi: 10.2196/jmir.1868] [Medline: 22569642]

Abbreviations

FUE: follicular unit extraction
SIM: subscriber identity module

Edited by G Eysenbach; submitted 29.04.13; peer-reviewed by D Horsfall, A Jabbaria, R Deonandan; comments to author 22.05.13; revised version received 28.12.13; accepted 19.01.14; published 06.02.14.

Please cite as:
Ozan-Rafferty ME, Johnson JA, Shah GH, Kursun A
In the Words of the Medical Tourist: An Analysis of Internet Narratives by Health Travelers to Turkey
J Med Internet Res 2014;16(2):e43
URL: http://www.jmir.org/2014/2/e43/
doi:10.2196/jmir.2694
PMID:24513565

Original Paper

Seeking Insights About Cycling Mood Disorders via Anonymized Search Logs

Elad Yom-Tov[1], PhD; Ryen W White[2], PhD; Eric Horvitz[2], MD, PhD

[1]Microsoft Research, Herzeliya, Israel

[2]Microsoft Research, Redmond, WA, United States

Corresponding Author:
Elad Yom-Tov, PhD
Microsoft Research
13 Shenkar St
Herzeliya, 46875
Israel
Phone: 972 747111359
Fax: 972 747111359
Email: eladyt@yahoo.com

Abstract

Background: Mood disorders affect a significant portion of the general population. Cycling mood disorders are characterized by intermittent episodes (or events) of the disease.

Objective: Using anonymized Web search logs, we identify a population of people with significant interest in mood stabilizing drugs (MSD) and seek evidence of mood swings in this population.

Methods: We extracted queries to the Microsoft Bing search engine made by 20,046 Web searchers over six months, separately explored searcher demographics using data from a large external panel of users, and sought supporting information from people with mood disorders via a survey. We analyzed changes in information needs over time relative to searches on MSD.

Results: Queries for MSD focused on side effects and their relation to the disease. We found evidence of significant changes in search behavior and interests coinciding with days that MSD queries are made. These include large increases (>100%) in the access of nutrition information, commercial information, and adult materials. A survey of patients diagnosed with mood disorders provided evidence that repeated queries on MSD may come with exacerbations of mood disorder. A classifier predicting the occurrence of such queries one day before they are observed obtains strong performance (AUC=0.78).

Conclusions: Observed patterns in search behavior align with known behaviors and those highlighted by survey respondents. These observations suggest that searchers showing intensive interest in MSD may be patients who have been prescribed these drugs. Given behavioral dynamics, we surmise that the days on which MSD queries are made may coincide with commencement of mania or depression. Although we do not have data on mood changes and whether users have been diagnosed with bipolar illness, we see evidence of cycling in people who show interest in MSD and further show that we can predict impending shifts in behavior and interest.

(J Med Internet Res 2014;16(2):e65) doi:10.2196/jmir.2664

KEYWORDS

information retrieval; mood disorders; bipolar disorder; machine learning

Introduction

People in the United States and other developed countries spend a significant portion of their time online [1]. Anonymized logs of such activities provide an unprecedented opportunity for studies in public health as well as applications that serve individuals in a private manner, helping them monitor and improve the quality of their lives. One recent survey shows that 81% of people in the United States use the Internet [2]. When they seek medical information, 59% of those people search for it online. Web search engines play an important role in the provision of medical information to health consumers, especially because of the anonymity of online search [3], allowing people to comfortably seek out and review sensitive information.

Search logs gathered by commercial search engines such as Google, Bing, and Yahoo! enable privacy-sensitive analyses of people's search behavior in the aggregate, across populations of users. The use of online resources for syndromic surveillance has been termed *infodemiology* [4]. Possibilities for leveraging signals from the Internet arise in the large-scale use of Web search for accessing health information. Prior research has studied multiple aspects of long-term search behavior in a medical context [5], for example, to identify flu outbreaks [6], improve medical retrieval [7], and to identify information needs of patients [8]. More generally, search log data has been used to study how people perform searches [9], in order to predict their next online actions [10,11], to predict their future interests [8], to improve search engines [12,13], and to understand in-world activities from long-term logs [5,14]

We present here analyses of the online behavior of people exhibiting an intense interest in mood stabilizing drugs (MSD), medications prescribed for helping patients with mood disorders. We find evidence that observed behaviors may be associated with episodes of mood swings and show that atypical periods of anomalous online behavior can be detected and predicted by observing past behaviors and comparing them to current observed behaviors.

Mood disorders (MD) are defined as a group of diagnoses in the Diagnostic and Statistical Manual of Mental Disorders (DSM IV) classification system relating to the changes in a person's affective state. Mood disorders affect a significant percentage of the population, though varying ranges of incidence (9.3%-23.3%) have been reported [15,16]. The commonality of these diagnoses is that a disturbance in mood is the main underlying feature. Cycling mood disorders are characterized by intermittent episodes of the disease. Bipolar disorder (BD) is characterized by intermittent episodes of shifts in mood across a spectrum of affect, including periods of mania or hypomania. Such episodes degrade the quality of life for those afflicted. Therapy and medications are employed to limit mood-swing episodes. Medications include antidepressants, antipsychotics, and mood stabilizers.

Previous work [8] has shown that users who seek specific health information (eg, specific disease names) are largely patients and health care professionals. Thus, we shall assume that many of those searching for MD medications are those likely suffering from MDs, and we offer supporting evidence for this conjecture in the results presented later in the article. We analyze these searchers' online behaviors and attempt to predict the onset and persistence of what appear to be episodes of significant swings in mood. We identify a candidate cohort by finding users with a high level of interest in specific mood stabilizing drugs (MSDs). Having identified these users and noted key aspects of their search behavior, we analyze a survey taken by people who have been diagnosed with a mood disorder. The data from the survey provides evidence that the searchers are more likely to be people suffering from mood disorders than they are health care professionals seeking information on patient care. The survey also provides evidence that shifts seen online (Dataset 1) with users' interests in particular topics (eg, in adult material consumption) are correlated with episodes of mania. Finally, we show how changes in information needs of these users may indicate forthcoming mood swings. This work highlights the possibility of constructing applications that could operate in the privacy of patients' own computing systems and serve to provide predictions about the likelihood of impending episodes. Such predictions might be used one day to guide preparations in advance of episodes or to gate interventions that might be used counter or minimize the level of debilitation associated with a mood swing.

Methods

We extracted all English language queries submitted to the Microsoft Bing search engine by users in the United States for the six-month period from December 2011 to May 2012 (inclusive). We refer to this dataset as Dataset 1. For each query, we extracted the query text, time and date, a list of pages visited by the user as a result of the query, and an anonymized user identifier. An anonymized user identifier (a string hash) was generated and stored in a Web browser cookie on the user machine, enabling logging of Bing search queries and clicks on search results for computers over time. No other data were available via this method. We note two intrinsic limitations of the data used in the study: (1) we cannot distinguish between multiple users on the same machine, and (2) if a searcher uses the search service on multiple devices, they would appear in logs with separate identifiers, one per device.

In order to maintain user privacy, data were first anonymized by hashing, before the investigators had access to them. They were then aggregated prior to analysis and no individual-level user datum was examined by the experimenters. The Microsoft Research Ethics Advisory Committee reviewed and approved the methods and results, and provided insightful discussion and guidance on the study.

We defined queries on mood stabilizing drugs (MSD queries) as those containing the following specific drug names: Eskalith, Lithobid, Lithonate, Lithotabs, Valproic acid, Divalproex, Valproate, Depakote, or Depakene, as well as queries specifically mentioning the term "mood stabilizing" (and its derivatives) or Lithium (except where the term was used in conjunction with the term "battery", "ion", and similar terms).

A total of 127,803 users made such queries during the period of study. We note that the market share of Bing was reported to be approximately 16% during the data period (see, for example, [17]), and so the number of users querying for drugs is only a sample of the total population taking MSD.

In order to focus on people who are likely using MSD, we used a threshold of 5 MSD queries during the study period (similar to [8]) to identify users with a high level of interest in MSD. We identified 20,046 such users and extracted all queries posted by these users during the study period. On average, each of the latter users submitted 34 queries per week on all topics.

We used a proprietary classifier developed by the Microsoft Bing team to assign each query into a set of 63 categories, including, for example, commerce, tourism, video games, weather-related, and adult-themed queries. The classifier is used by Bing to determine whether to display special results such as instant answers. Queries could be classified into multiple categories (eg, purchase of flight tickets would be classified into both tourism and commerce).

To validate findings derived from analysis of the data drawn from Bing search, we performed identical analyses on behavioral data collected from an opt-in consumer panel recruited by Internet analytics company comScore. Millions of panelists provide comScore with explicit permission to passively measure all of their online activities using monitoring software installed on their computers. In exchange for joining the panel, participants are offered a variety of benefits, including computer security software, Internet data storage, virus scanning, and chances to win cash or prizes. In addition to logged search behavior, the comScore data also provides us with panelists' gender and age (mostly bucketed in 5-year increments). Overall, the panelists were 53.04% (45,707/86,168) female, with the most prevalent age range being 25-34 years. We refer to the comScore data as Dataset 2.

Beyond the studies of online logs, we conducted a survey among 272 people who self-identified as being prescribed one of the drugs listed above. Respondents were recruited using the online survey website "Instant.ly". The survey was comprised of 11 multiple-choice questions and 9 free-text questions. The survey is provided in Multimedia Appendix 1. We refer to the survey data as Dataset 3.

Results

Search Behaviors

A first set of results were generated on Dataset 1 via analysis of the searches from computers where the threshold number of 5 or more MSD queries were observed. We find that the pattern of observed searches changes significantly around the time that queries on MSD are issued. We provide evidence that such observed shifts in behavior are linked to the onset of mood swing episodes. After providing evidence that MSD queries can be used as a label for the onset of such events, we present a second set of results on the feasibility of predicting forthcoming episodes.

Are Users Identified Through MSD Queries Likely to be Suffering From Mood Disorders?

Some users state the purpose of their query, for example, using queries such as "I have severe depression". We counted the number of unique queries that mention a mood disorder (either using the term or using one of the terms "depression", "mania/manic", or "bipolar") and whether or not they appeared in conjunction with the person to whom the mood disorder refers (either "I" or "my wife/ husband/ spouse/ son/ daughter/ boyfriend/ girlfriend"). Queries in the first person that mentioned mood disorders were 6.5 times (659 compared to 102) more likely than queries about mood referring to other people (statistically significant, *P*<.001, chi-square test). We take this as evidence that queries on mood disorders without reference to self or other tend to refer to the searcher.

Some of the above-mentioned drugs are also prescribed to people suffering from seizures or migraines. We found a ratio of 2.23 between the number of who people mentioned a term associated with mood disorders (the term itself or "depression", "mania/manic", or "bipolar") and queried for MSD and the number of users who mentioned migraines or seizure and also queried about MSD drugs versus those who mentioned a term at the threshold numbers. This validates our focusing on users with mood disorders by analyzing the drugs prescribed for these disorders.

Online Behavior Before and After MSD Queries

On average, queries on MSD were entered by searchers every 13 days (SD 11). While most users only rarely posted such queries, a sizable population of searchers posted MSD queries multiple times and on a large number of days during our study period. Users who made MSD queries, on average, made 17.6 queries per day on days when MSD queries were posted, compared to 8.7 queries (SD 16 for both) on days when they were not (statistically significant, *P*<.001, sign test).

For 98.42% (14,410/14,641) of MSD queries mentioning a specific drug name, a single drug was mentioned. A single drug was queried during the study period by 70.31% (5010/7126) of users who mentioned a specific drug name. Even among users who posted 10 or more queries containing an MSD name, 97.9% (860/878) mentioned a single drug name. Thus, user interest is predominantly in a single drug, even when interest in the drug is high.

Figure 1 shows the distribution of queries by time of day and by day of the week, partitioned by MSD queries and all other queries made by users in Dataset 1. As displayed in the figure, there is a slight increase in MSD queries during morning hours and on weekdays (vs weekends) (statistically significant, two-sample Kolmogorov-Smirnov, *P*<.001). This aligns with known MD patterns, where the most severe symptoms typically occur in the morning [18] and changes in daily routines are known to induce mood episodes [19].

Using the method described in [14], we synchronized the query stream of users to each day they posted an MSD query, and for each day measured the probability of posting queries in each of the query classes, compared to all queries. For the population studied, the categories of *nutrition*, *business*, and *adult* showed more than a doubling of the query probability as compared to all other query categories. Figure 2 shows the change in the likelihood of querying in these categories. The jump in query probabilities peaked at 3.30 for nutrition queries, 2.29 for business queries, and 2.24 for adult queries. We will show later that those who self-reported being afflicted with mood disorders reported similar increases in related topics when they experienced a mood swing. Comparing behavior before and after the time of the MSD query (marked as time zero), we note that all three categories show a large increase in likelihood on the day after the MSD query. This is especially noticeable in the adult queries, which show a diurnal cyclic behavior that increases in magnitude until four days after the MSD query. Interestingly, the peak in these queries is in the morning hours (8am-9am), corresponding to the peak of diurnal activity of MSD queries.

Figure 1. Probability of posting MSD queries as a function of time of day (above) and day of the week (below), compared to all other queries. This figure shows that MSD queries are more common in morning hours of weekdays.

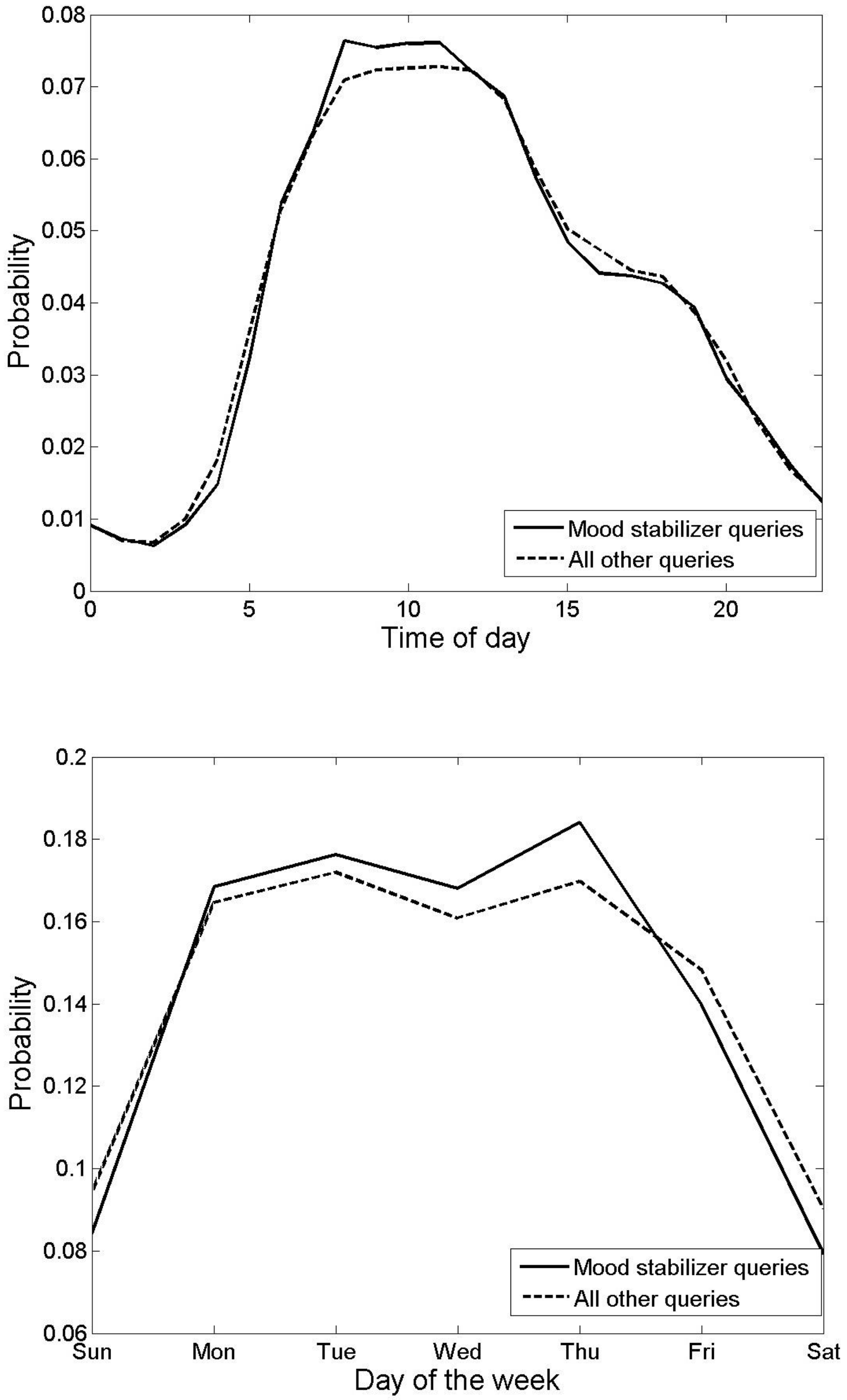
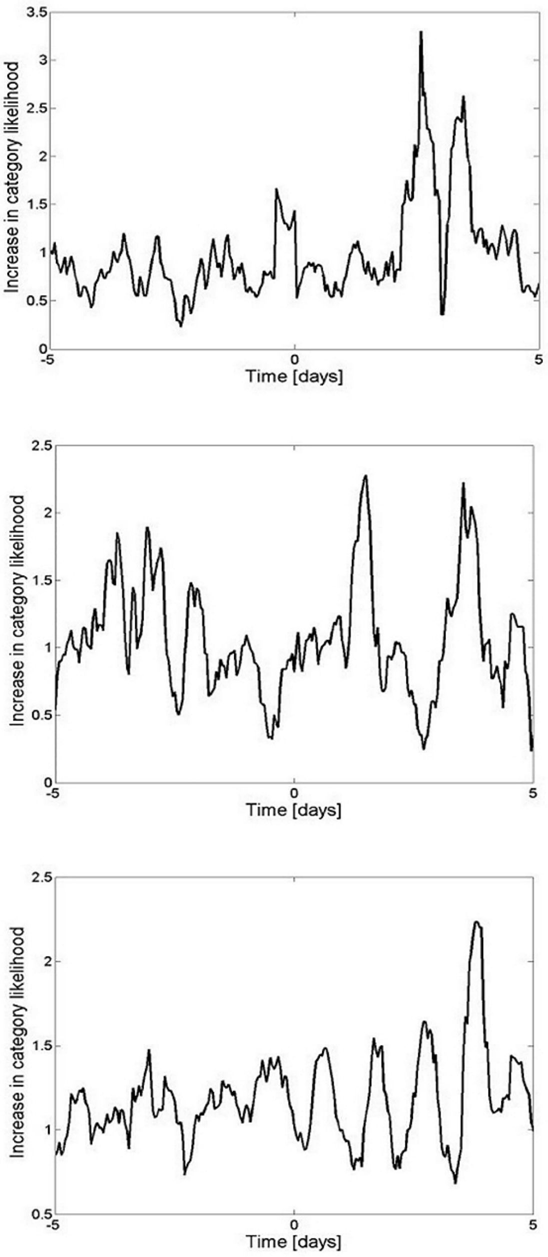


Figure 2. Change in query category likelihood as a function of time. Zero marks time of an MSD query. Categories represented are (from top to bottom): nutrition, business, and adult materials. The time series was smoothed with a moving average 5 hours in length.



Characterizing MSD (Mood Stabilizing Drug) Queries

The most common terms in MSD queries, excluding stop words and the drug names themselves, were manually partitioned by us into four mutually exclusive categories (with percentages of users mentioning terms in category in parentheses; see Table 1.

One of the most common concerns of people entering MSD queries are medication side effects. This is most noticeable when observing people who post MSD queries on at least 5 days of the study period. For these users, adverse drug reactions are an ongoing concern and the most common terms were “side effects” (both as a term pair and each word separately).

Table 1. Common MSD query terms.

Search term category	% of users mentioning terms (n=20,046) n (%)	Terms
Side effects	5391 (26.89%)	mood, effects, side, side effects, weight, effect
Drug-related	1847 (9.21%)	mg, long, take, dosage, dose, generic, release, interaction, gain
Disease-related	1516 (7.56%)	cause, bipolar, depression, high, low, treatment, anxiety, symptoms
Other	2467 (12.31%)	drug, blood, levels, children, use, sod, medication, code, sprinkles, used, taking, loss, time, list, normal, test, work, help, liver, patient, pain, lab, together

Characterizing Clicked Pages

The informational goals of people querying on terms used to refer to MSD may be ambiguous. For example, a query for “lithium” could refer to the drug or the metal. To address such ambiguity, we analyzed the search results that users had clicked on using the click logs from the Bing search engine over the period of the study (Dataset 1). Queries followed by clicking on relevant links that are titled and summarized with text snippets are less likely to be ambiguous.

With MSD queries, we count the number of users and clicks that lead to each Web domain such as drugs.com. We remove effects of multiple clicks (eg, which may reflect user learning during the session) by analyzing the first result clicks for each query only. To remove noise from the click data, we focus only on clicks followed by a long dwell (30 or more seconds) on the landing page. Long dwells have been shown to correlate with satisfaction in prior studies of information-seeking behavior [20]. Table 2 lists the top 10 most popular domains, ranked in descending order by the number of users.

The results reveal that the top 10 clicked domains are health-related with the exception of wikipedia.org, answers.yahoo.com, and wiki.answers.com, which are comprised of reference information and social question answering data. We note that the average number of clicks (considering one click per session) per user is greater than 1, indicating that people visit these sites multiple times. We note that some sites (lower in the list) are unrelated to MSD and are linked to alternate meanings of “lithium” (for example, a channel on satellite radio, a song by Nirvana, a fitness training website, batteries, and mining companies). Given the prevalence and ambiguity of “lithium”, we removed or isolated users who queried only for this term in our later analysis.

Table 2. Top 10 clicked URLs following MD medication queries.

Clicked domain	No. of users	No. of clicks	Avg. no. clicks / user
drugs.com	5688	9989	1.756
en.wikipedia.org	4711	7136	1.515
ehow.com	2493	3252	1.304
wiki.answers.com	2270	3607	1.589
answers.yahoo.com	2199	3024	1.375
bipolar-disorder.emedtv.com	1786	2159	1.209
webmd.com	1529	1853	1.212
ncbi.nlm.nih.gov	1326	1827	1.378
healthcentral.com	1294	1540	1.190
bipolar.about.com	1237	1489	1.204

Searcher Demographics

We next analyzed search logs purchased from comScore (Dataset 2), comprising user search behavior from comScore panelists over a 12-month timespan that overlapped with the period of time covering Dataset 1. Unlike Dataset 1, the comScore logs provide demographic information about searchers. The comScore data includes searches issued to all major Web search engines (Google, Bing, and Yahoo!), providing a broader user sample than the set of Bing users in Dataset 1. We also examined users who queried for at least one of the specific drug names mentioned above. In these data, users were tracked using an anonymous identifier connected to each individual panelist rather than with a machine-based identifier as in Dataset 1, which does not allow us to discriminate among multiple users of a single machine.

Beyond moving from potentially multi-user machines to individual searchers, the comScore data also provides searchers’ age ranges and gender. We sought to understand whether the distribution of searchers exhibited demographics that match those of known MSD patients. The dataset includes 1116 users who queried for MSD medications. A tornado diagram covering age and gender for a control group of 100,000 randomly chosen panelists is shown in Figure 3. Another tornado diagram displayed in Figure 4 shows the distribution of age and gender for users who query for MSD drugs. In comparing the test and the control group, we can see that women are more likely than men to query for MSD medications (control: 53.04%, 45,707/86,168, female vs MD users: 57.06%, 586/1027, female, statistically significant, two-proportion *z* test, *P*=.005). Although prevalence of bipolar disorder is similar across genders, the difference may reflect the fact that women are more likely to experience more severe effects and bipolar II disorder than men [21]. We also note a spike in queries on MSD, in the 25-34 age range for men and women, which is higher for men (statistically significant, chi-square test, *P*<.001). This aligns well with studies of mood disorders, which have shown that the median age for the development of mood disorders is 30 years [22]. These results provide evidence that the user group observed searching for MSD medications online may be similar demographically to the subset of the population known to be affected by mood disorders.

Figure 3. Age and gender distribution across a control group of 100,000 users.

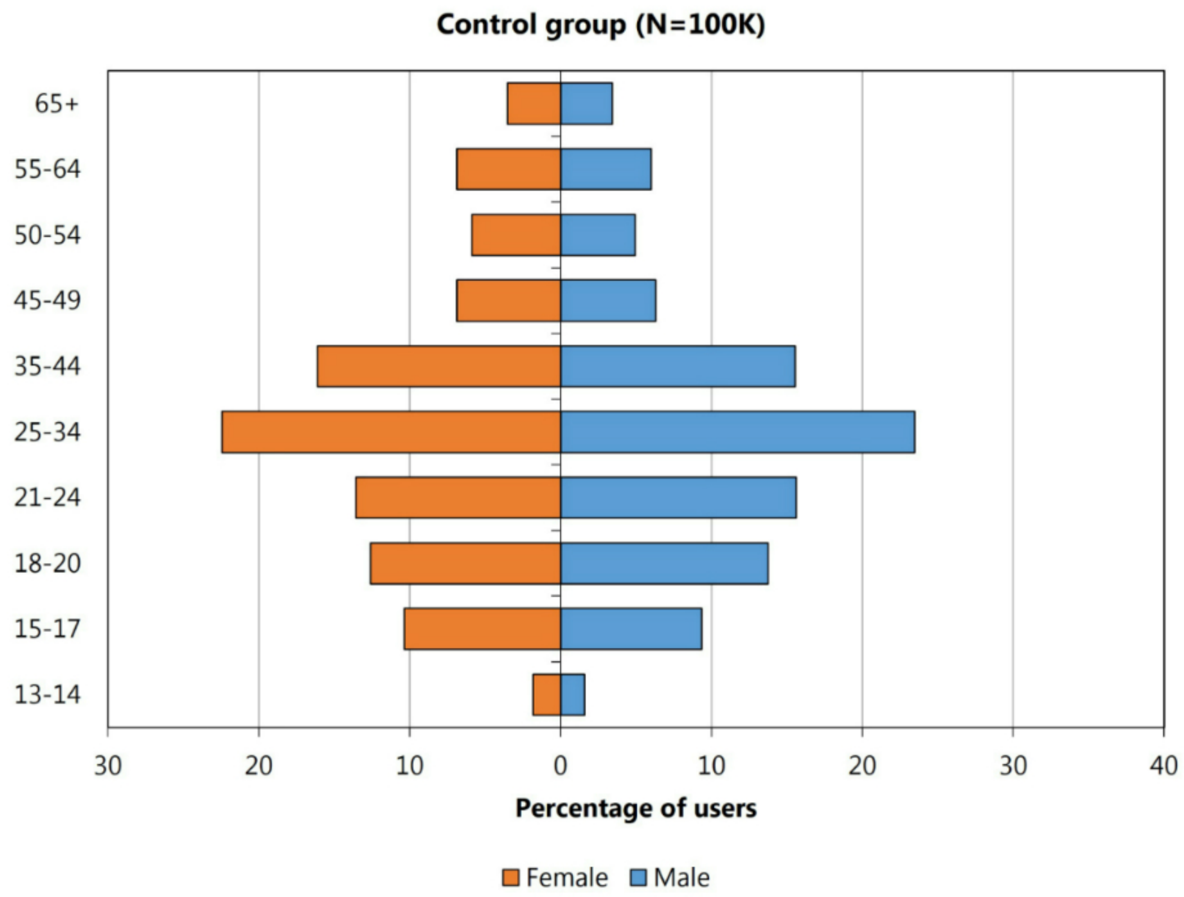
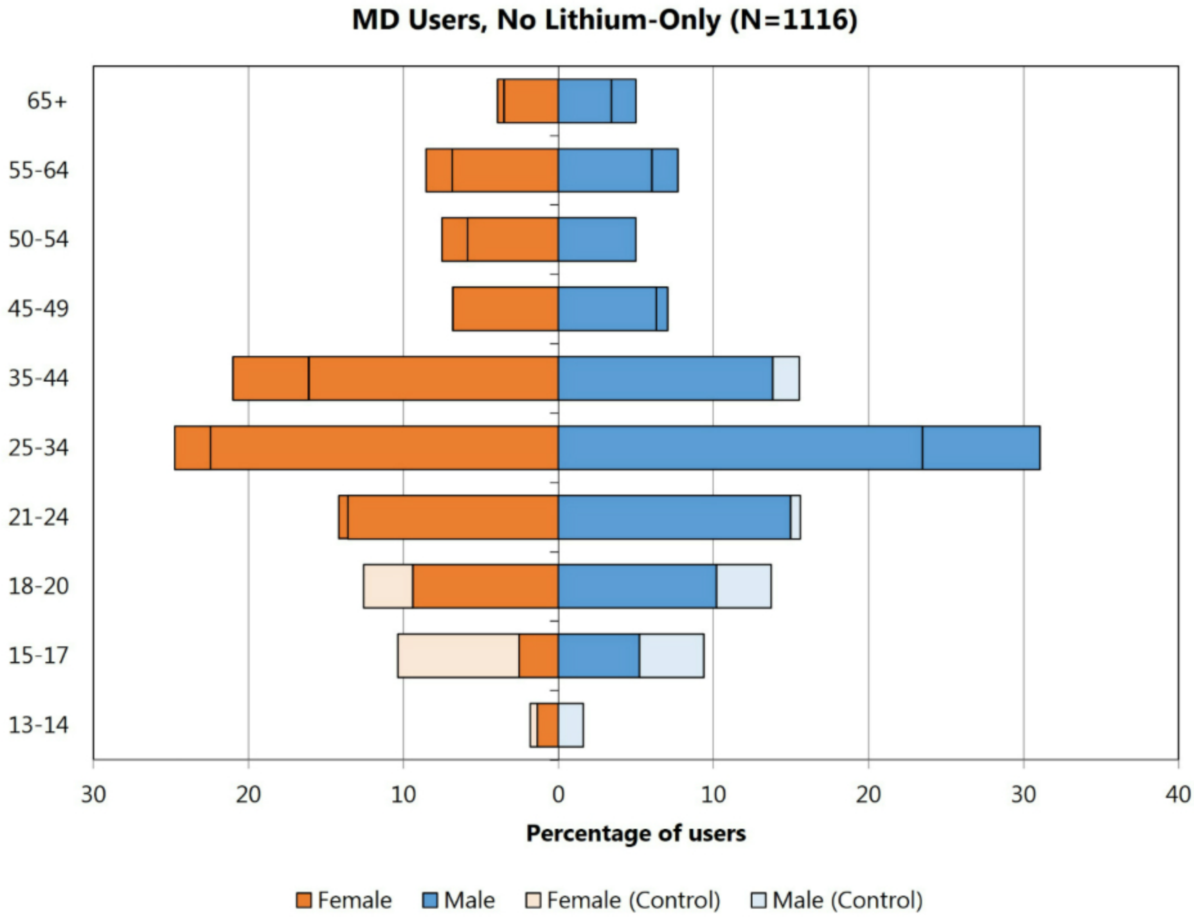


Figure 4. Age/gender distribution of users who searched for mood disorder medications, minus users who only searched for lithium. MSD plot is overlaid on the control plot from Figure 3. When the control value is larger, the control bars (in lighter shades) are visible. When the MSD value is larger, the MSD bars occlude the controls and are marked with lines in the MSD bar.



Survey Data

To more fully understand the search behavior observed in the logs in Datasets 1 and 2, we conducted a complementary survey of 272 people who had self-identified as being prescribed one of the MSD listed above (Dataset 3). People in the survey cohort had a median age of 36 years (min: 18, max: 77), 25.4%, 56/220, were male, and had taken an MSD for an average of 3.2 years. Respondents reported having an average of 10.5 episodes per year, lasting an average of 5.5 hours. Approximately 97.7% (256/262) of respondents were prescribed a regular dose of the drug, 71.7% (188/262) daily, 17.9% (47/262) more than once daily, and the rest at less than daily doses), and 83.6% (214/256) reported that they comply with their prescription.

Only 18.3% (48/262) of respondents never searched for information on their MSD on the Internet; 30.9% (81/262) reported searching once in the past 6 months, 44.6% (117/262) between twice and 10 times, and 6.1% (16/262) reported searching more than 10 times. Thus, approximately 51.9% (133/262) of respondents made repeated queries about their medications. The most common triggers for searching were the first time that the drug was prescribed (71.8%, 188/262), when side effects occur (42.7%, 112/262), when the respondent felt the drug was not working (32.4%, 85/262), following discussions with friends or family (29.4%, 77/262), and at the outset of an episode (25.2%, 66/262).

Respondents sought information on side effects (82.1%, 215/262), efficacy (62.2%, 163/262), dosage (40.5%, 106/262), and retail locations for acquiring the medication (13.0%, 34/262). They found information at reference websites such as Wikipedia (55.3%, 145/262), consumer-oriented sites (47.3%, 124/262), social media (40.6%, 106/262), and information posted by drug manufacturers (38.2%, 100/262). These findings, particularly on the frequency searches for side effect and dosage information and the types of resources selected, align well with the log analysis performed on Dataset 1.

People who searched for information multiple times reported doing so because they needed more information (42.7%, 79/185), wanted reassurance that the drug was the right one for them (27.6%, 51/185), needed different information than before (18.9%, 35/185), or because they had forgotten information they had once known (10.2%, 19/185).

Approximately half (48.6%, 70/144) of the respondents reported that they changed their online search and browsing behavior when they experience a manic (“high”) state, and a similar percentage (55.9%, 79/179) reported a change in behavior during depression (a “low”). Of the respondents that provided input, 8.5% (41/144) mentioned doing unnecessary online shopping during manic states, 22.2% (32/144) reported researching new topics on the Web, 13.9% (20/144) look for information that will make them happy (including their hobbies), 13.9% (20/144) reported being more active on the Web in general, 8.3% (12/144) search for health information, and 6.2% (9/144) reported an elevated interest in searching sites with pornographic content.

Respondents reported that, during depression periods, they usually stay offline (68.2%, 122/179), a finding that is in line with those reported by De Choudhury et al [23]. Others reported looking for depressing topics (15.1%, 27/179) or shopping, so as to feel bad (2.8%, 5/179), which resonates with their depression. Others looked for health information (7.3%, 13/179) or sites that could make them happy (6.1%, 11/179).

Predicting Queries via Users' Query Streams

We posit that most of the searchers posting MSD queries above the threshold rate are actual patients taking MSDs and that the MSD information seeking is likely performed near the commencement of an MD event, possibly the start of a manic episode. The former is evidenced by the temporal patterns associated with such postings (ie, days of the week and time of day), the fact that most users were interested in a single drug, and the demographics of searchers. The latter is apparent from the behavioral changes associated with days before and after MSD queries are posted. For example, we see a jump in business-related (eg, shopping) queries (see Figure 2), an activity that was identified in the survey as correlated with manias. Moreover, a quarter of survey respondents reported searching for MSD when they feel that an episode is commencing, which suggests that some proportion of recurrent MSD queries are triggered by the onset of a mood swing. We believe that most MSD queries are related to manic events because, as reported in the survey, people who experience depression tend to stay offline.

Given that the days on which MSD queries are posted may be significant, we focus on the prediction of days on which MSD queries will be posted. We investigated three distinct populations of users: (1) Recurring MSD, (2) Occasional MSD, and (3) Lithium Only. Recurring MSD included users who posted MSD queries on at least 5 days during the data period. There were 498 users in this population. Occasional MSD were users who posted an MSD query that mentioned a specific MSD drug in a query, but did so on fewer than 5 days. There were 9633 users in this population. Lithium Only users posted a query that mentioned Lithium, but not a specific drug, and did so on fewer than 5 days. There were 9884 users in this population.

We represented the users' daily query streams as vectors of the attributes detailed in Table 3. We augmented the daily activity vector of a user by concatenating the vector with the average vector of activity for up to the last 14 days (ie, the average of the past two days, the average of the past three days, etc.) Also, for each attribute in Table 3, we computed the difference between the attribute (eg, the number of queries in category) and average value of that attribute in all previous non-MSD days. These attributes represent divergence between typical and current activity. Finally, the day of the week was added as an attribute. Thus, a total of 1981 attributes (14 lags + 1 divergence, times 132 attributes, and one day of the week attribute) were used to represent the daily activity of each user.

We used the feature representation of the user to predict whether the user would post at least one MSD query in the following day. A separate classifier was trained for each user population. In order to obtain a valid comparison among populations, we chose a random subset of the Occasional MSD and Lithium Only populations of the same size as the Recurring MSD population (so as to afford similar sized training datasets), and report results on those subsets.

We constructed a decision tree [24] as a classifier. Five-fold cross-validation [24] at the level of searchers (to avoid problems of training on future data) was used to train and test the classifier; each user was randomly assigned to one of the five cross-validation folds. Classifier performance was measured by identifying the area under the receiver operating characteristic curve (AUC).

The AUC for the three populations is shown in Figure 5. The highest AUC is obtained for the Recurring MSD population and the lowest for the Lithium Only population. The difference between the two MSD populations is not statistically significant, but differences are significant between the Lithium Only population and the MSD populations ($P<.001$) [25]. The findings could be attributed to differences in the number of samples per user or to the number of positive labels, although no statistically significant correlation was found. It is likely that the differences are in some way connected to inherent differences among the user populations.

The attributes selected more than twice among the 50 highest levels of the decision trees are listed in Table 4. First, we note that the number of attributes chosen for the MSD populations are much fewer than those selected for the Lithium Only population, suggesting that the first two populations are more homogeneous than the latter population. Second, we note the prominent appearance of adult-related queries in the Recurring MSD population, which is likely related to the activities shown in Figure 2. We also note that many of the attributes appearing in the MSD populations (and to a lesser extent, the Lithium Only population) are related to hobbies, a fact underscored by the reported interests of survey participants.

In view of the observed spike in adult-related queries following MSD queries, we evaluated the use of adult-related queries as an outcome and label and sought to construct predictive models to forecast whether a searcher would ask an adult-related query on the next day. We focused on a population that (as in the Recurring MSD population) asked an adult-related query in at least 5 days of the study period. There were 275 such users, 5.8% (16/275) of which also appeared in the Recurring MSD population and 77.1% (212/275) in the Occasional MSD population. The AUC for this prediction task was 0.71 (compared to 0.78 for MSD queries), suggesting that adult material is a strong proxy for behavioral changes followed by mood stabilizing events, an effect noted previously in several studies [26,27].

Figure 5. Area under the ROC (AUC) for the three populations of users.

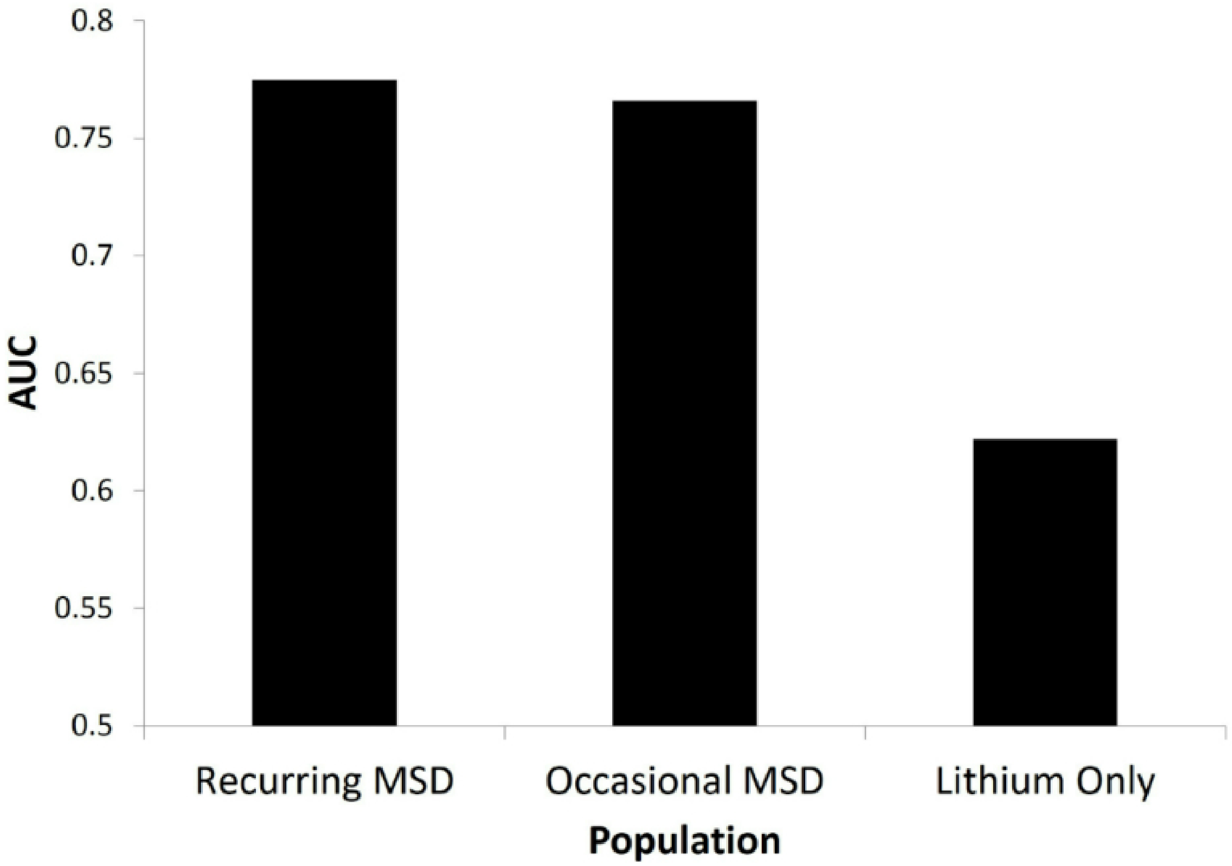


Table 3. Attributes of classifier for predicting days when mood stabilizing drug (MSD) queries are posted.[a]

Attribute	Number of attributes
Total number of queries per day	1
Total number of query topics per day, as represented by categories identified by query classifier.	1
Maximal number of queries per hour	1
Number of active hours per day	1
Number and fraction of queries posted in unusual hours (defined as 11pm to 4am local time).	2
Number of queries from each category (both in raw scores and after thresholding)	126

[a] The attributes are computed for each of 14 time lags, as well as the divergence from non-MSD activity, for a total of 1980 attributes.

Table 4. Most frequently selected attributes for each class of users.

User population	Features
Frequent MSD [a]	
	Number of adult queries
	Divergence in the number of adult queries
	Number of vehicle-related queries
	Number of commerce-related queries
	Divergence in the number of commerce queries
	Number of event-related queries
	Divergence in the number of queries related to flight status
Occasional MSD	
	Day of the week
	Number of vehicle-related queries
	Number of book-related queries
	Divergence in the number of commerce queries
	Divergence in the number of celebrity-related queries
	Number of queries related to clothes and shoes
	Number of commerce-related queries
	Divergence in the number of commerce queries
Lithium Only	
	Day of the week
	Number of vehicle-related queries
	Divergence in the number of vehicle-related queries
	Number of book-related queries
	Divergence in the number of book-related queries
	Number of commerce queries
	Number of queries related to clothes and shoes
	Divergence in the number of commerce queries
	Divergence in the number consumer electronics queries
	Number of event-related queries
	Divergence in the number of finance-related queries
	Number of queries related to flight status
	Number of health-related queries
	Divergence in the number of health-related queries

[a] MSD: mood stabilizing drug

Discussion

Principal Findings

Analyses of logs of search activity show potential as a valuable tool in public health as well as for privately fielded applications and services that work on behalf of users. We identified a population of users who show a strong interest in MSD and also show significant changes in their online search behavior around the time of expressing interest in the medication. We showed that we can build predictive models that can be used to forecast the future appearance of MSD-centric search queries, which may be associated with the onset of a bipolar episode. We believe the possibility of predicting mood swing episodes with applications running within the privacy of a user's own computing device might one day help patients and caregivers to better understand and prepare for impending changes in mood.

Comparing the results of the survey with the behavior observed online, we find several similarities. First, the ratio of the number of people who made between 2 and 10 queries and those who made more than 10 queries compared to those who made a single query is 1.4 and 0.2 in the survey, respectively, compared to 0.5 and 0.1 in the query log data. Second, the topics of search identified by respondents closely match those of the frequently occurring terms on MSD queries. Finally, respondents reported searching for information related to shopping, hobbies, and health information, as well as heightened search on adult content, which correspond to our observations on the online behavioral dynamics coinciding with users seeking information about MSD.

The appearance of repeated queries for MSD is an intriguing phenomenon. Survey respondents reported re-submitting queries on MSDs because they needed more or different information, or because they needed reassurance that the prescribed drug was the right one for them. Web data suggests that the latter rationale, together with concerns about side effects, were major causes for such repeat searches. We posit that the onset of a mood swing episode causes patients to become more aware of their disease in general and of their drugs in particular (especially when they are non-compliant), triggering an MSD search. Thus, such triggering behavior may not be limited to mood disorders, but to a wider class of diseases.

Limitations

A key limitation of our study is the lack of a gold standard on outcomes. We cannot directly link users in our cohort with their real-life persona, and thus cannot know if they are, in fact, suffering from a bipolar mood disorder. To address this shortcoming, we note that previous work [8] has shown that seeking specific health information is done mostly by patients and primary caregivers. Second, the observed change in online behavior around the time of MSD queries provides evidence that users are experiencing a major change in their activities during this period of time. Third, the temporal appearance of MSD queries (at specific times and dates) and the demographic profile of these searchers are consistent with actual patients taking MSDs. Finally, users self-reported in a survey that they make similar changes in their online behavior. The most discriminatory features in our predictive models were those that quantified the change in behavior in several specific areas, including health and commerce. These topic areas were independently identified by survey respondents (in response to free-text questions) as topics that receive attention during manic periods. These findings provide evidence in support of our association of MSD queries with mood swing episodes and provides evidence that we are observing users who experience these events, especially of the mania type.

Conclusions

Although drugs and behavioral treatment can reduce the incidence of mood disorder events, compliance with MSD prescriptions has been estimated to be as low as 35% [28]. The low compliance may be based on the side effects of these drugs [29]. However, non-compliance is associated with greater severity of mania events [30]. Predicting MD events before they occur may help people to better understand and prepare for changes in mood. We believe our study sets the stage for improved health and well-being for patients with mood disorders.

Opportunities for future work include working with patients to link online activities with clinical observations. Such efforts could validate our results and ascertain the accuracy of early warnings, as well as verify whether timely forecasts about an impending episode could be harnessed in beneficial ways.

Authors' Contributions

All three authors conceived the idea for the study and developed the methodology for its execution. EYT extracted and analyzed query data. RWW extracted and analyzed comScore data. All authors designed the survey and EYT analyzed it. All three authors were involved in final analysis and writing of the manuscript.

Conflicts of Interest

The authors are employees of Microsoft Research. All work was done as part of the respective authors’ research, with no additional or external funding.

Multimedia Appendix 1

Survey of users taking MSDs.

[PDF File (Adobe PDF File), 96KB - jmir_v16i2e65_app1.pdf]

References

1. Sverdlov G. Forrester. 2012. Understanding the changing needs of US online consumers URL: http://www.forrester.com/Understanding+The+Changing+Needs+Of+US+Online+Consumers+2012/fulltext/-/E-RES84361?docid=84361 [accessed 2014-02-12] [WebCite Cache ID 6NKP8yiZv]
2. Fox S, Duggan M. Pew Internet & American Life Project. 2013. Health Online URL: http://www.pewinternet.org/Reports/2013/Health-online.aspx [accessed 2014-02-12] [WebCite Cache ID 6NKPIE7C7]
3. Pelleg D, Yom-Tov E, Maarek Y. Can you believe an anonymous contributor? On truthfulness in Yahoo! Answers. 2012 Presented at: ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust; 3-5 Sept. 2012; Amsterdam, The Netherlands p. 411-420. [doi: 10.1109/SocialCom-PASSAT.2012.13]

4. 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]
5. White RW, Horvitz E. Studies on the onset and persistence of medical concerns in search logs. 2012 Presented at: ACM SIGIR conference on Research and Development in Information Retrieval; 2012; Portland, OR, USA p. 265-274. [doi: 10.1145/2348283.2348322]
6. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
7. Rubin DL, Flanders A, Kim W, Siddiqui KM, Kahn CE. Ontology-assisted analysis of Web queries to determine the knowledge radiologists seek. J Digit Imaging 2011 Feb;24(1):160-164 [FREE Full text] [doi: 10.1007/s10278-010-9289-2] [Medline: 20354755]
8. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of information-seeking for cancer on the Internet: an analysis of real world data. PLoS One 2012;7(9):e45921 [FREE Full text] [doi: 10.1371/journal.pone.0045921] [Medline: 23029317]
9. White RW, Drucker SM. Investigating behavioral variability in web search. 2007 Presented at: World Wide Web; 2007; Banff p. 21-30. [doi: 10.1145/1242572.1242576]
10. Lau T, Horvitz E. Patterns of search: Analyzing and modeling web query refinement. 1999 Presented at: User Modeling; 1999; USA p. 119-128.
11. Downey D, Dumais S, Horvitz E. Models of searching and browsing: languages, studies, and applications. 2007 Presented at: International Joint Conference on Artifical intelligence; 2007; Hyderabad.
12. Dupret G, Piwowarski N. A user browsing model to predict search engine click data from past observations. 2008 Presented at: ACM SIGIR conference on Research and Development in Information Retrieval; 2008; Singapore p. 331-338. [doi: 10.1145/1390334.1390392]
13. Joachims T. Optimizing search engines using clickthrough data. 2002 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2002; Edmonton p. 133-142. [doi: 10.1145/775047.775067]
14. Richardson M. Learning about the world through long-term query logs. ACM Trans. Web 2008 Oct 01;2(4):1-27. [doi: 10.1145/1409220.1409224]
15. Bijl RV, Ravelli A, van Zessen G. Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Soc Psychiatry Psychiatr Epidemiol 1998 Dec;33(12):587-595. [Medline: 9857791]
16. Grant BF, Stinson FS, Dawson DA, Chou SP, Dufour MC, Compton W, et al. Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Arch Gen Psychiatry 2004 Aug;61(8):807-816. [doi: 10.1001/archpsyc.61.8.807] [Medline: 15289279]
17. comScore. Dec. 2012 US search engine rankings. URL: http://www.comscore.com/Insights/Press_Releases/2013/1/comScore_Releases_December_2012_U.S._Search_Engine_Rankings [accessed 2014-02-14] [WebCite Cache ID 6NNrKKtFK]
18. McClung CA. Circadian genes, rhythms and the biology of mood disorders. Pharmacol Ther 2007 May;114(2):222-232 [FREE Full text] [doi: 10.1016/j.pharmthera.2007.02.003] [Medline: 17395264]
19. Frank E, Gonzalez JM, Fagiolini A. The importance of routine for preventing recurrence in bipolar disorder. Am J Psychiatry 2006 Jun;163(6):981-985. [doi: 10.1176/appi.ajp.163.6.981] [Medline: 16741196]
20. Fox S, Karnawat K, Mydland M, Dumais S, White T. Evaluating implicit measures to improve web search. ACM Trans. Inf. Syst 2005 Apr 01;23(2):147-168. [doi: 10.1145/1059981.1059982]
21. Ketter TA. Diagnostic features, prevalence, and impact of bipolar disorder. J Clin Psychiatry 2010 Jun;71(6):e14. [doi: 10.4088/JCP.8125tx11c] [Medline: 20573324]
22. The numbers count: Mental disorders in America.: National Institute of Mental Health URL: http://www.nimh.nih.gov/health/publications/the-numbers-count-mental-disorders-in-america/index.shtml [accessed 2014-02-14] [WebCite Cache ID 6NNsBGfd1]
23. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2013 Presented at: AAAI Conference on Weblogs and Social Media; 2013; Boston.
24. Duda RO, Hart PE, Strok DG. Pattern Classification and Scene Analysis 2nd ed. In: Pattern classification. New York: John Wiley & sons; 2001.
25. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982 Apr;143(1):29-36. [Medline: 7063747]
26. Långström N, Hanson RK. High rates of sexual behavior in the general population: correlates and predictors. Arch Sex Behav 2006 Feb;35(1):37-52. [doi: 10.1007/s10508-006-8993-y] [Medline: 16502152]
27. Stein DJ, Black DW, Shapira NA, Spitzer RL. Hypersexual disorder and preoccupation with Internet pornography. Am J Psychiatry 2001 Oct;158(10):1590-1594. [Medline: 11578986]
28. Li J, McCombs JS, Stimmel GL. Cost of treating bipolar disorder in the California Medicaid (Medi-Cal) program. J Affect Disord 2002 Sep;71(1-3):131-139. [Medline: 12167509]

29. Scott J, Pope M. Nonadherence with mood stabilizers: prevalence and predictors. J Clin Psychiatry 2002 May;63(5):384-390. [Medline: 12019661]
30. Keck PE, McElroy SL, Strakowski SM, Stanton SP, Kizer DL, Balistreri TM, et al. Factors associated with pharmacologic noncompliance in patients with mania. J Clin Psychiatry 1996 Jul;57(7):292-297. [Medline: 8666570]

Abbreviations

AUC: area under the receiver operating characteristic curve
MD: mood disorder
MSD: mood stabilizing drug

Edited by G Eysenbach; submitted 10.04.13; peer-reviewed by N Smalheiser, A Parks, H Mueller, S Antani; comments to author 15.06.13; revised version received 30.09.13; accepted 10.01.14; published 25.02.14.

Please cite as:
Yom-Tov E, White RW, Horvitz E
Seeking Insights About Cycling Mood Disorders via Anonymized Search Logs
J Med Internet Res 2014;16(2):e65
URL: http://www.jmir.org/2014/2/e65/
doi:10.2196/jmir.2664
PMID:24568936

Original Paper

Why Do Patients and Caregivers Seek Answers From the Internet and Online Lung Specialists? A Qualitative Study

Romane Milia Schook[1], MD; Cilia Linssen[2], MSc; Franz MNH Schramel[3], MD, PhD; Jan Festen[2], MD, PhD; Ernst Lammers[4], MD; Egbert F Smit[1], MD, PhD; Pieter E Postmus[1], MD, PhD; Marjan J Westerman[5], PhD

[1]VU University Medical Center, Department of Pulmonary Diseases, Amsterdam, Netherlands
[2]Lung Cancer Information Center: Longkanker Informatiecentrum, Maarsbergen, Netherlands
[3]St Antonius Hospital Nieuwegein, Department of Pulmonary Diseases, Nieuwegein, Netherlands
[4]Gelre Ziekenhuizen, Zutphen, Netherlands
[5]Vrije Universiteit Amsterdam, Department of Health Sciences, Methodology and Applied Biostatistics, Amsterdam, Netherlands

Corresponding Author:
Romane Milia Schook, MD
VU University Medical Center
Department of Pulmonary Diseases
De Boelelaan 1117
Amsterdam, 1081 HV
Netherlands
Phone: 31 20 4442193
Fax: 31 20 4444328
Email: r.schook@vumc.nl

Abstract

Background: Since its launch in 2003, the Dutch Lung Cancer Information Center’s (DLIC) website has become increasingly popular. The most popular page of the website is the section “Ask the Physician”, where visitors can ask an online lung specialist questions anonymously and receive an answer quickly. Most questions were not only asked by lung cancer patients but also by their informal caregivers. Most questions concerned specific information about lung cancer.

Objective: Our goal was to explore the reasons why lung cancer patients and caregivers search the Internet for information and ask online lung specialists questions on the DLIC’s interactive page, “Ask the Physician”, rather than consulting with their own specialist.

Methods: This research consisted of a qualitative study with semistructured telephone interviews about medical information-seeking behavior (eg, information needs, reasons for querying online specialists). The sample comprised 5 lung cancer patients and 20 caregivers who posed a question on the interactive page of the DLIC website.

Results: Respondents used the Internet and the DLIC website to look for lung cancer–related information (general/specific to their personal situation) and to cope with cancer. They tried to achieve a better understanding of the information given by their own specialist and wanted to be prepared for the treatment trajectory and disease course. This mode of information supply helped them cope and gave them emotional support. The interactive webpage was also used as a second opinion. The absence of face-to-face contact made respondents feel freer to ask for any kind of information. By being able to pose a question instantly and receiving a relatively quick reply from the online specialist to urgent questions, respondents felt an easing of their anxiety as they did not have to wait until the next consultation with their own specialist.

Conclusions: The DLIC website with its interactive page is a valuable complementary mode of information supply and supportive care for lung cancer patients and caregivers.

(J Med Internet Res 2014;16(2):e37) doi:10.2196/jmir.2842

KEYWORDS

lung cancer; patients; caregivers; website; online lung specialists; reasons; Internet; information needs; coping; qualitative

Introduction

Lung cancer is the second most common type of cancer and the most common cause of cancer deaths in both men and women in Europe and the United States [1,2]. The Netherlands counted approximately 14,000 lung cancer patients in 2002. This number increased to more than 21,000 in 2011 [3]. In 2003, the Dutch Lung Cancer Information Center (DLIC) was established. Its purpose was to give simple, accessible, and evidence-based information on lung cancer through its website [4], as well as support, and to unite lung cancer patients and their caregivers [5]. The unique quality of the DLIC is that it was set up at the national level and broadly supported by both health care professionals and patient groups. Following its creation, a special section was added to its website—an interactive webpage, called “Ask the physician”, where visitors could anonymously ask an online lung specialist questions and receive an answer within 48 hours [6]. It was after some scepticism from the lung specialists involved in the website’s management and content that this new section was launched [6].

Since its launch, the DLIC website has become increasingly popular and has reached a steady number of 25,000 unique visitors per month [5-7]. Surprisingly, the most popular page of the website is the “Ask the Physician” page. Our study group has previously investigated who was visiting the interactive webpage and what the information visitors were looking for [6,8]. Seventy-four percent of the questions (n=1893) were not only asked by lung cancer patients (13%) but also in large numbers by their informal caregivers (61%, eg, family, friends, and loved ones). Most questions (43%) concerned specific and general information about lung cancer. Furthermore, verification of information given by patients’ own specialists was sought, for example, the overall survival rate of lung cancer or specific therapeutic advice.

This impressive number of caregivers visiting the DLIC website, next to lung cancer patients, and their use of the online experts piqued our interest. Literature shows that a cancer diagnosis is an immediate reason for an increase in Internet use by patients and their families to obtain medical information, next to unmet information needs [9-13]. Looking for information seems to be an important and frequent task for caregivers, next to other activities, such as patient care, supporting and facilitating medical consultations, and aiding in information recall [14-16]. Also, physicians have limited consultation time and are not statutorily obliged to address or inform caregivers as they are patients, although such an approach is recommended [17,18]. Furthermore, consultations addressing multiple persons require high communication skills and are thus difficult. It is therefore plausible to think that unmet information needs underlie the above mentioned Internet use. However, if we look at the total picture, many elements and motives for these Internet searches remain unclear to us, especially with regard to caregivers of lung cancer patients in particular and the patients themselves. Why are there so many caregivers of lung cancer patients looking for information and consulting the DLIC online lung specialists? Why do they address the online specialists instead of the patient’s own specialist? Are there differences between caregivers and patients with regard to their motives when seeking information and their needs?

In comparison to other groups of patients, lung cancer patients and their caregivers are more vulnerable because the patients are facing a much shortened life expectancy. They need more special attention from health care providers. As the quality of life, psychological adjustment to the disease, risk of appraisal, anxiety, and depression of cancer patients and their caregivers are affected by barriers and failures in information supply and communication from health care providers [19-21], it is crucial that this vulnerable group receive information meeting their needs, especially because these needs differ throughout the cancer trajectory [22].

Adequate information supply is considered to be a part of good medical care and support. To provide appropriate care to lung cancer patients and their caregivers, it is important to explore their information-seeking behavior in order to gain more insight into their needs and indications for better communication modes and a tailored information supply.

Therefore, we conducted a qualitative interview study with telephone interviews to explore the reasons why caregivers and lung cancer patients search the Internet for information and ask the DLIC online lung specialists additional questions about lung cancer, next to face-to-face consultations with their own treating specialists.

Methods

Design

We conducted a qualitative, exploratory study consisting of semistructured telephone interviews with visitors to the DLIC website who asked the online lung specialist a question on the “Ask the Physician” webpage. This study was approved by the local medical ethics committee of the VU University Medical Center, Amsterdam.

Procedures and Participants

Recruitment took place between August 2005 and April 2006. All consecutive visitors who asked the online lung specialists a question on the interactive page of the DLIC website were invited to participate in our study. After submission of a question, an online (digital) confirmation window would pop up with Dutch text, containing an invitation for participation in our study and an explanation about its purpose and the telephone interviews to be held. If visitors wished to participate, they were asked to complete an online form with their name, phone number, and home and email address in order to be contacted. After completion of the form, visitors could click on the button “send”. If they did not wish to participate, they could just close the pop-up window.

Within 3 weeks, visitors who had given their online consent for study participation were contacted by email or phone by CL (a communications expert and expert interviewer with no prior relationship to the study respondents). CL gave them additional information and checked whether participants were fluent Dutch speakers. When CL contacted the volunteers by phone and oral participation consent was given, they were either interviewed immediately, or a later appointment was made. If CL contacted them by email, written participation consent was given and an appointment was made for a future telephone interview.

Of the 84 persons who had agreed to participation online, 43 individuals could not be reached, 4 asked to postpone the interview but never contacted CL again, and 7 refused participation after initially having been interested (4 were due to the patient's death/current poor condition, and 3 gave no reason). Ultimately, 30 participants were interviewed on their motives for looking for information on the Internet and asking questions on the DLIC "Ask the physician" webpage. Of the 30 participants, 5 were patients with a lung tumor, 20 were caregivers of lung cancer patients, and 5 did not have cancer (one lung patient and 4 individuals who feared that they had lung cancer).

Since we were interested only in cancer patients with a lung tumor and their caregivers, we analyzed only the 25 respondents with cancer. The ratio of patients and caregivers (5:20) is not balanced but is in accordance with the population of visitors of the interactive webpage, as we have reported in other papers [6,8]. One patient had small cell lung cancer (SCLC), three had non-small cell lung cancer (NSCLC), and one probably had breast cancer with lung metastases. The patients had a median age of 52 years (range 44-62). The majority of caregivers were women, most of them daughters and partners, with a median age of 39 years (range 21-58) (see Table 1 for more details on participants).

Table 1. Study population characteristics and Internet use (n=25).

	Gender M/F	Age, years	Diagnosis patient[b]	Previous Internet use[b]	Education[c]	Caregiver type	Current therapy
P[a]1	F	44	Metastasized BC	No	HE	–	Palliative therapy
P2	M	62	NSCLC I/II	Yes	HE	–	Adjuvant chemo
P3	F	50	SCLC ED	Yes	LE	–	Palliative chemo
P4	M	52	NSCLC SU	Yes	LE	–	Palliative chemo
P5	F	52	NSCLC I/II	Yes	HE	–	Adjuvant chemo
CG[a]1	F	57	NSCLC IV	Yes	HE	Partner	Palliative therapy
CG2	F	36	NSCLC IV	Yes	LE	Daughter	Deceased 6 weeks before
CG3	F	45	NSCLC IV	Yes	HE	Partner	Deceased 3 months before
CG4	F	52	NSCLC I/II	Yes	HE	Partner	After surgery, no adjuvant chemo
CG5	F	39	Mesothelioma	Yes	HE	Daughter	Terminal phase
CG6	F	39	SCLC LD	No	HE	Daughter	No current therapy, chemoradiation 1 year before
CG7	F	33	LC SU	Yes	LE	Daughter	Palliative chemo
CG8	F	51	NSCLC IV	Yes	LE	Partner	No current therapy, palliative chemo 6 months before
CG9	F	32	LC SU	Yes	LE	Daughter	After diagnostics and diagnosis
CG10	F	26	LC SU	Yes	HE	Niece	Deceased
CG11	M	22	Mesothelioma	Yes	LE	Nephew	Therapy unknown 6 months after diagnosis
CG12	F	42	LC IV	N/A	LE	Daughter	Palliative therapy
CG13	M	58	LC IV	Yes	LE	Partner	Palliative therapy (radiotherapy)
CG14	F	38	LC IV	Yes	LE	Sister	Palliative therapy
CG15	F	21	NSCLC I/II	Yes	HE	Daughter	No current therapy, surgery 1 year before
CG16	M	28	SCLC SU	Yes	HE	Son	Therapy unknown 3 months after diagnosis
CG17	M	36	LC IV	Yes	HE	Partner	Palliative chemo
CG18	M	35	LC IV	N/A	HE	Son	Palliative therapy
CG19	F	41	Mesothelioma	Yes	HE	Daughter	Deceased recently
CG20	F	44	LC SU/ metastases BC	Yes	LE	Daughter	Therapy unknown during diagnostics

[a] P=patient, CG=caregiver.

[b] BC=breast cancer, NSCLC=non–small cell lung cancer, IV=stage IV, SCLC=small cell lung cancer, ED=extensive disease, SU=stage unknown, I/II=stage I or II, LD=limited disease, LC=lung cancer type unknown.

[c] HE=high education (university, academy, college level), LE=low education (primary school, high school, intermediate vocational training).

Interviews

A topic list was made to prepare the interviews. Main topics were Internet use, information needs and supply, reasons to use the Internet, reasons to query the online lung specialists on the DLIC website, and reasons not to. The topic list was completed with personal information about demographics and disease. Sample questions asked during the interviews are listed in Textbox 1.

The semistructured telephone interviews were conducted by CL. Participants were encouraged to talk freely until all topics were discussed. The duration range of an interview was 20-90 minutes. All interviews were written down verbatim with pen and paper, put into orthographic transcripts, and then subsequently typed into MS Word documents directly after interview termination (these MS Word documents will be referred to as “interview transcripts”). During transcription into MS Word documents, CL would already start to classify interview passages according to their content and the questions listed in Textbox 1 (topic list based). Apart from the interview transcripts, sometimes CL wrote notes with her impressions on the respondents’ ideas during the interviews, which she attached as a memo to the interview transcripts. Occasionally she also copied interesting quotes from her email correspondence with the participants as field notes into the interview transcripts.

Textbox 1. Sample interview questions.

When did you start looking for information on the Internet?

Why did you look for information on the Internet?

What role does the Internet play in information supply?

What role does the caregiver play in information supply?

Why did you ask a question on the interactive page of the DLIC website?

What did you ask? What did you want to know?

Was the answer to your question satisfactory? Was it useful? Why?

At which moment during the lung cancer procedure did you have the greatest information needs?

Why did you not ask the (patient’s) treating physician your question?

Is it different to ask a question through the Internet/by email? Why? How so?

What is your opinion about the possibility of asking an online physician questions on a website?

What is your opinion about the possibility of asking your treating physician questions by email?

What is your opinion about the possibility of asking a nurse questions by email?

Would you like to communicate with the (patient’s) treating physician by email?

What would be the value of such communication?

Analysis

Researchers RMS and MJW used a thematic approach in the analysis of the transcripts (n=25) [23]. After familiarization with the data by reading it repeatedly and carefully, we made a summary of each interview and started initial coding of the transcripts. To facilitate coding, organizing, collecting, and selecting data from the interview transcripts, we used MaxQDA version 10 [24]. After numerous meetings focusing on understanding the collected data and correct interpretation, we determined potential themes first and then sorted and collated the (initial) codes according to them. Hence, we looked at the participant’s Internet use first, to further determine when they started to surf the Internet and to assess their information needs. After this, we focused on identifying and classifying the reasons why participants surfed the Internet and posed questions to online lung specialists instead of their own specialist. Grouping this information, we made an initial thematic map (available on request).

After review of the potential themes for coherence, we refined these themes, identified new themes, and recoded some data extracts. This refinement led to the identification of similarities and discrepancies between participants with regards to the sought-after information. The newly identified themes were found to be the beneficial effects of looking for information for participants, the presence of tension between patients and caregivers provoked by the Internet search, and perspectives about the use of email with the patient’s specialist. After recoding the data extracts according to the refined and new themes, we reviewed the entire dataset again and discussed the generated main themes conscientiously and critically for coherence, consistency, robustness, and representativeness [23,25] in order to develop a final thematic map (available on request).

Results

Starting to Surf the Internet

All respondents, except for 1 patient and 3 caregivers, reported using the Internet on a regular basis. They had access to Internet at home and used it for daily activities such as checking their email, banking, or looking for different types of information. They reported that the lung cancer diagnosis specifically urged them to seek information and ask the online lung experts lung cancer–related questions. Their diagnosis had a great impact on their lives, as they were facing lung cancer, its (future) treatment trajectory, and ultimately the shortened life expectancy of the patient. Therefore, they felt the need for additional information to learn how to deal with the situation by any means.

Both patients and caregivers also mentioned that they surfed the Internet again at specific moments later during the lung cancer treatment trajectory, such as during chemotherapy, at the appearance of new symptoms or disease progression, or when having to make a choice between two treatment options. These moments also meant a change in their current vulnerable balance, which pushed them to search for information again.

> *I have been told a lot at the hospital, but everything goes so fast, you hear a lot of terms, and you just do not know anything […] First, I looked at the tumor types and how everything would go during surgery. After that, I looked again when it was recommended for me to have chemotherapy.* [P2]

Patients and caregivers mentioned that their need to seek information often arose once they had time to rest and think about what they had been told, often at a time when their questions could not directly be answered by the treating specialist anymore: “Once you have come home, you have forgotten half of what you have been told, which is exactly the moment you would want to ask something.” [CG8, partner]

What Are Respondents Looking For and Why?

Respondents reported searching for lung cancer–related information in general but also information specific to their personal situation. An illustration of the information search of caregiver Sylvia (fictitious name), describing what she was looking for and why, is given in Textbox 2.

Apart from feeling the need to gather general information in order to be better informed and have a better understanding of the disease, respondents wanted to be prepared for future consultations, future course of disease, and treatment trajectory. They also felt the need for specific information regarding practical matters or emotional support directly related to their individual condition in order to help face current or short-term situations. Examples are practical matters during the treatment trajectory and finding emotional support through contact with fellow sufferers (see Table 2). Ultimately, all respondents expressed that the main goal of their information search was to find support as they were dealing with lung cancer. Textbox 3 also gives an illustration of patient Mary’s (fictive name) information search and her search motives.

Textbox 2. Caregiver Sylvia.

Sylvia (fictitious name) is the 36-year-old daughter of a lower educated male patient of 72 years old. At time of the interview, Sylvia’s father had died a week earlier. He had been diagnosed with metastasized lung cancer 7 weeks prior.

> *I started looking [for information] 2 weeks after the definitive diagnosis of lung cancer with brain metastases had been made. I searched the Internet because I wanted to know the prognosis and what different types of lung cancer there were. Once I started, I kept on going. I also wrote something on the forum of the DLIC website and I got some reactions; it was very nice. It may sound strange, but it is nice to know that there are many people who are dealing with the same thing. At the hospital they don’t have much time for you. You can see them thinking “Yes, you have cancer, I have explained everything, now get on with it”. Then you come home and the questions arise […] and you think “I want to ask the question now”. But if you call the doctor, you get the secretary who says “the doctor is not here, he is with a patient. When do you have an appointment? Next week? You can ask your questions then”. But this way a question that feels urgent to you remains unanswered. This is one of the reasons why I turned to the Internet and by stumbling across the DLIC website, […] I found everything I was looking for. This website is incredible. I can stay on the site for hours […] When I came back to the treating physician after the diagnosis, I asked him “what kind of lung cancer is it, small cell, non–small cell? What are the advantages and disadvantages of giving therapy?” You should have seen his face wondering how I knew all of that. Actually, I have only ever asked the online expert one question: “if someone has metastasized cancer, why is not it possible to operate on the lungs and brain and just remove the cancer from both sites?” I received a satisfactory answer. Although it was just as I thought, it was still nice to get confirmation. And you never know, perhaps the Internet expert will say there are still possibilities or new therapies. Even if it is not the case, it is still nice to have been able to ask. I think it is excellent to be able to ask a question of the DLIC online expert and to get an answer so quickly because it has prevented several sleepless nights.*

Textbox 3. Patient Mary.

Mary (fictitious name) was a 44-year-old highly educated, married patient. She was diagnosed with breast cancer metastasized to the lungs. After a period of stable regression, she was receiving palliative therapy at time of the interview. She was very pleased with Internet as mode of information supply and the DLIC interactive webpage, but she emphasized that she did not want eHealth to become a substitute for visual contact with treating specialists in the future.

> *After the diagnosis, I hit rock bottom. A neighbor, who is a nurse, brought me a lot of pamphlets. This helped me back on top and gave me the feeling that I should do something. At that point [1 month after the diagnosis], I started to study the folders and the Internet. I wanted to come to grips with the situation and also get the feeling that I had a rough plan for my treatment. The Internet has played an important role in terms of information supply. I am a member of a private mailing group where we exchange a lot of information. One of the group members made me aware of the DLIC website and that questions were being answered there. Early on, I would see my doctor first and after that I*

> *would go online looking for the things he had suggested, verifying whether there were no other possibilities. But there came a point when I felt I needed to take charge of the situation instead of just following him passively. I wanted to get ahead of the game, so now I started looking for information before every new hospital appointment, so that I could come well prepared. I researched every possible thing: what I had, what I would get, what possible alternative or experimental therapies there were, things about pain control for later during the process, you name it. I have accepted the disease and I see myself as a kind of research project. I think it is good to be well informed. It gives you the feeling that you have control over your disease and your treatment.*

Table 2. Needs of respondents.

Information needs	Examples
General information needs	
Background information on lung cancer	Epidemiology, lung cancer types
Diagnosis	
Therapy	Regular, experimental, alternative
Diagnostics	Investigations types and explanation about it
Disease course/end of life/prognosis	Life expectancy, (overall) survival per stage, what to expect at the end of life, suffocation
Information sources/literature	DLIC website, websites specialized on experimental therapy
Individual information needs	
Managing personal situation	Help with a choice: postoperative chemotherapy, radiotherapy or not
Managing personal health or mental condition	Explanation and/or treatment of symptoms or side effects (eg, own neurological problems or insensibility after surgery), preparation for coming treatment (eg, what is going to happen during surgery), analgesia, what can this symptom be?
Managing emotions	Search for hope, confirmation, reassurance, emotional support, compassion, consolation, contact with fellow sufferers/ comparable experiences, expert's verification/2ndopinion, moments of panics and uncertainty
Managing daily life	Lifestyle advices, hospital bills, food supplements, hospice
Managing practical aspects of treatment trajectory	Eg, explanation of medical terms/terminology like stable disease or "adenocarcinoma", meaning of laboratory/imaging results specific to patient, organizing euthanasia

Differences Between Patients and Caregivers

Although patients and caregivers reported searching for general information as well as information specific to the individual condition, their searches differed with regard to quantity and content (see Textboxes 2 and 3). Patients searched for a minimal amount of general background information and focused specifically on their individual current condition. They aimed to get more information about symptoms and therapy, together with practical information for the coming disease course and consultations. Not all of the patients seemed to want to gather information about the last moments of life, although they were perfectly aware of their shortened life expectancy. Patients especially said they were searching (among other things) for hope by contacting fellow sufferers and/or looking for a confirmation of the accuracy of their diagnosis and chosen management. This made them feel supported (see Textbox 3).

On the contrary, caregivers expressed the need to collect a lot of general information of any kind (see Textbox 2). An element they mentioned with regard to the available information on the Internet was the difficulty of understanding or interpreting online information correctly, as they were lacking a doctor's knowledge and felt overwhelmed by the vast amount of information given. Caregivers also wanted to be informed more frequently about the patients' end of life and prognosis in particular. Furthermore, they reported looking for information specific to their personal situation in order to feel supported. Like the patients, caregivers said that their quest for (emotional) support consisted of (among others) searching for hope and reassurance/confirmation of medical information, contact with other fellow sufferers (by direct contact or by reading their stories) and with the online expert (see Table 2).

Why Not Ask Their Own Treating Specialist?

Numerous reasons were mentioned for using the Internet and asking the online expert questions in particular, rather than addressing their own treating specialist. Respondents said they did not *want* to ask their own specialist. They mentioned being ashamed about discussing personal matters or indecencies within the context of a consultation. They also did not wish to disturb or burden their own specialist because he or she was (supposed to be) very busy. Especially caregivers were convinced of being a burden to the specialist as they were "only a caregiver anyway".

> *There are a lot of things that I wouldn't discuss with the lung specialist. For example, a big part of his lung has been removed and when we caress each other, he has no feeling in that part of his skin. He does not*

> *feel my touch. He says that it feels alien, as though it's not part of his own body. Then we asked ourselves "will it remain that way?" But you do not ask the lung specialist those kind of things when he is looking at the chest X-ray very seriously.* [CG4, partner]
>
> *And you do not bother the specialist by calling him at the hospital, because surely he has better things to do.* [P3]

Respondents had the feeling of not being able to ask a question or request information. They experienced barriers in their contact with their own specialist because they felt he or she had no time during and between consultations. They experienced the specialist not being open to questions: "the communication at the hospital was dramatically bad" and "We have a kind of reticence to ask our treating specialist questions. They do not like it when you ask a question." [CG6, child]

> *Following the suggestion of the DLIC, I brought a list of questions for my specialist. You could see him thinking "Not another one with a list…" Specialists are always in a hurry. They do not even have time for a proper discussion with you. I have seen 4 different pulmonologists, and when you walk into the consultation room, they would all still be reading your medical record at the same time.* [P2]

Patients as well as caregivers appreciated the convenience of using the Internet and the DLIC website because of its 24-hour accessibility and its anonymity. This made them feel freer to ask the online lung experts for any kind of information and helped them express their feelings better. They felt less anxiety as they were able to pose a question instantly, receive a reply from the online specialist to urgent questions within a short space of time, and not have to wait until the next consultation with their treating specialist. This was especially the case for caregivers as, for example, one of them said that the patient had a follow-up appointment every 3 months but that she could surf the Internet every day. Furthermore, respondents appreciated the expertise and open-minded, kind, and empathic attitude of the online lung experts.

> *I think it's fantastic. It is anonymous and it's great to ask your questions to someone who's competent in the field […] I was looking for someone independent […] Although he [the online expert] is an outsider, he knows what he is talking about.* [P2]
>
> *I think it is very good to be able to ask a physician questions online. It's a smaller step to take than calling or talking to your treating specialist.* [CG9, child]
>
> *His style (of the online expert) is really nice, not disapproving. He is very kind and always says something like "I wish you good luck" or "I hope it will be all right".* [CG1]

Reasons Not to Use the Internet

Patients and caregivers mentioned that sometimes they postponed or stopped their Internet search, for instance, because the information they encountered was too much. Not searching helped them to stay positive. Other respondents felt that they had collected enough information after a period of time and therefore deliberately quit the information search, knowing they were avoiding confrontation with the disease sometimes.

> *No, at the beginning, I did not look for information. The disease, it was not about me, it was as if it was about someone else […] I must say that I am not on the Internet very often anymore, because a lot of people die there. It is too much for me [...] and makes me feel depressed. You need to feel there is still a light at the end of the tunnel.* [P3]
>
> *I went looking for information after my husband passed away, not during his illness though. It does not make you feel happy and I wanted to stay positive, so it does not help if you read these unpleasant things on the Internet.* [CG3, partner]
>
> *I think I should not read too much about lung cancer anymore. Now that the disease has gone, it's time to move on. I have got the feeling that I have just recovered from a heavy illness myself and that at long last I am finally fit enough to get up and go again. Yes, perhaps I am avoiding thinking about all that is lost.* [P2]

Tensions

Both patients and caregivers also talked about the occurrence of tension when meeting their own information needs by searching the Internet. Specifically caregivers realized that their needs were not always the same as the patients' and experienced difficulties in dealing with the information they had collected. They felt torn by the dilemma of disclosing sensitive information or hiding it from the patients, as they wanted to protect them from (unwelcome) confrontations. For example, one caregiver said that he did not share the death of someone from his mailing group as he thought that this would be too much to handle for the patient.

> *There are things that I do not tell him, because I do not want to worry him […] It is difficult because sometimes, when we are at the doctor's, I would like to know things, such as the life expectancy, but I am reluctant to ask, because I do not want my father to hear it.* [CG12]
>
> *Well…Actually I have kept information from him when he was very unwell and we did not know yet whether he could be operated. At the time, my son and I looked for the 5-year survival rate and decided we should not tell him. Once he was home again after the operation, he was looking at a very old medical encyclopedia that we never use. Then he said "Do you know what the 5 years survival of lung cancer is?" I said I did and in reply he asked me why I had never told him. He was upset at first, but he understood.* [CG4, partner]

Email Contact With Their Own Treating Specialist or an Oncology Nurse

All patients were very positive when being explicitly asked about their opinion on the opportunity to have email contact with their own specialist for questions. Caregivers, however, had more reservations as they felt embarrassed contacting the treating specialist (as being *only* caregivers) and were afraid that the specialists might be overwhelmed by emails.

Respondents also reacted positively about having email contact with an oncology nurse in order to obtain medical information and ask questions, on the condition that she or he had to specialize in lung cancer. One caregiver mentioned the very useful assistance of an oncology nurse as a constant and accessible contact point for support and information during the whole treatment trajectory.

Discussion

Principal Findings

The present study adds knowledge to the information-seeking behavior of lung cancer patients and their caregivers during the lung cancer treatment trajectory and their reasons for doing so. Strikingly, the majority of respondents were caregivers. Our findings show the coping strategies of caregivers and patients towards managing lung cancer. They searched the Internet and asked online DLIC lung experts questions because they wanted lung cancer-related information and help in coping with the disease practically and emotionally. This happened repeatedly during the whole treatment trajectory. This search helped the respondents to deal with lung cancer in a better way. It permitted them to gain a better understanding, be prepared (for the treatment trajectory and the disease course), feel free to search and ask for information, express feelings, be relieved of anxiety, feel emotionally supported, and regain control. This confirms that information is essential and beneficial for coping with cancer for both patients and caregivers and that caregivers are actively involved in information search and supply [14,26]. Furthermore, the information needs of caregivers differed from those of patients.

Lung cancer patients and their caregivers searched the Internet in order to deal with lung cancer and their personal situation. The perspective of the coping theory can be applied to explain the respondents' behavior [27]. When events occur in a subject's life, the subject is prompted to activate internal processes necessary to accommodate that event (eg, behavioral, cognitive, and affective mechanisms, including coping). It is known that anxiety, anger, fear, helplessness, and depressive feelings are frequently experienced after a cancer diagnosis [28]. The study respondents experienced such distress at diagnosis and other key points of the treatment trajectory that provoked a change in their lives, therefore posing a threat, challenge, or harm to them personally. Subsequently they tried to manage this distress by means of coping, through searching the Internet and turning to online lung experts for lung cancer-related information.

Information is essential for coping with cancer, and new media (eg, the Internet, online cancer communities, mailing groups, etc) are crucial today for the dissemination of information [14,26]. Hence, our respondents are used to searching new media for information. Still, the choice of the Internet and online experts versus the patient's treating specialist remains intriguing. Caregivers particularly had a greater tendency to search the Internet. This may be related to the serious nature of lung cancer, as the gravity of a disease urges people to seek additional information [14]. Also, when facing a life-threatening disease, cancer patients and family members often want confirmation of information, despite good communication with health care providers and adequate information supply [29,30]. Furthermore, Ong et al [31] and other investigators [32-35] found that patients and caregivers are often unsatisfied with the communication or the information given to them in medical settings. These issues were also observed in present study results. Apart from this, the practical advantages of the Internet and the availability and attitude of the online experts moved the respondents toward this medium.

Respondents were not only looking for general lung cancer-related information but also information specific to their own situation. Soothill et al [36,37] reported the need for "universal" and "personal" information among cancer patients and caregivers, helping them to cope with cancer, such as managing daily life or emotions. Searching for these two types of information was beneficial for the study respondents. Although similarities in the information search of patients and caregivers were observed, important differences were noticed too. Caregivers were inclined to look more extensively (in terms of quantity) for information than patients, and the content of the information found differed too. This trend was recognizable from the literature. Caregivers, of lung cancer patients in particular, show high participation rates in online cancer communities [14]. Compared to patients, they also have a higher tendency to look for information than providing it to other caregivers and patients, and they are more inclined to participate in emotional support exchange [6,8,14]. Moreover, lung cancer patients and caregivers have different information needs [38]; caregivers tend to have more unmet needs and concerns than patients [36,37]. This could originate from the caregivers' perception of themselves as being helpless observers, their lesser involvement with health care providers, or the patients' underreporting of concerns and unmet needs [39]. It also seems that information seeking is a typical activity for caregivers, as lung cancer patients are often too ill to do it themselves [14]. Interestingly, most caregivers among our respondents were (young) women. Women typically participate more in mail groups and supportive communication than men and seem to search or care more about information (provision) than men [14,40,41].

Thus, it is important to recognize the caregivers' needs as well as those of patient, since caregivers play a critical role in sustaining the cancer patient, and their ability to nurture and support the patient may be compromised in case of unmet needs [42,43]. This may have serious implications for both the patients' and caregivers' psychological state and coping. Further investigations on this topic are therefore needed.

As seen in our results, trying to meet one's own information needs can also be accompanied by difficulties and/or tensions. Confrontation with threatening or negative disease information can be of great impact on the well-being of patients and caregivers [11,44]. This may subsequently lead to the total abortion of the information search, the avoidance of confrontation with the "sensitive" information in particular, or concealing it from loved ones or patients, with all its possible consequences on the psychological state of those involved. This dilemma between wanting to meet information needs and protecting oneself or another have often been described in literature as the origin of conflicts and communication problems between caregivers and patients [45-47]. A balance between these two elements must therefore be achieved to maintain psychological well-being. Solutions to reach such a balance may not only be ceasing the information search temporarily, but also consulting reliable and clearly categorized sources of information and discussing the tension with the concerned persons or with someone who might be of help. The possibility of obtaining or discussing information directly with the treating physician or a specialized oncology nurse should be considered.

Study Limitations

Since we performed a cross-sectional study with an interview at one single point during the lung cancer treatment trajectory, it is possible that we have missed information on the respondents' needs, as we did not follow them over time. Nonetheless, respondents described different moments during lung cancer treatment. The quality of a person's information needs is constant over time even though the quantity of the needed information may show a slight decrease [48,49]. It is therefore reasonable to assume that our results paint a reliable picture of the information needs and other reasons why respondents surfed the Internet.

Another limitation is the fact that all interviews were held by phone and were not audio recorded. Also, CL started to classify interview passages shortly after their transcription. This may have led to bias and to information loss during the simultaneous transcription of the interviews. However, since CL is an expert interviewer used to collecting information in this particular manner, it seems less probable that data were lost. Furthermore, the classification of interview passages took place according to the questions CL asked during the interviews (see Textbox 1), reducing bias.

Because data collection took place a few years ago, changes in habits of Internet users and DLIC website visitors may have occurred over time, next to changes in website availabilities. The relevance of our findings may also be questionable. However, we know that Internet health searches have become much more commonplace [50]. Additionally, internal reports of the DLIC website have shown that the number of visitors each month and the visitors' identity remained the same over time [7]. The interactive webpage still remains a very popular page of the website, and the number of questions is still increasing. Questions concerning general information on lung cancer as well as information on personal matters are still being asked. The website availabilities have not really changed, and its homepage shows only a few additions since its launch (eg, animation, links to new blogs, and a visitor's poll). We can therefore assume that the reported findings are still relevant and representative for the population we investigated.

A final question to address is whether our study sample is representative of the investigated population, as ultimately a sample of 25 respondents was interviewed despite the far larger number of persons interested. Persons who never surf the Internet were also anticipatorily excluded. The respondents' distribution is, however, in accordance with the population visiting the DLIC website [6,8]. Moreover, patients and caregivers who never surf the Internet were not the target of this present study. Additionally, we showed in previous studies that many website visitors only read the webpage "Ask the physician" without asking questions. This group of visitors may be represented by the group of respondents who were interested in participating but were never interviewed. Another argument for the respondents' representativeness is the reaching of saturation of data and themes after multiple readings of the interviews.

Practical Implications

Our study results have numerous practical implications for the care of patients and their caregivers during the lung cancer treatment trajectory. Caregivers represented the majority of respondents looking for information and indicated they needed help coping with lung cancer. However, they often felt unable to address the patient's treating specialist. Since the well-being of patients and caregivers are connected, special emphasis must be given to the often neglected experience and needs of caregivers [51,52]. In practice, this might simply be solved by addressing caregivers' needs during consultations. In case of difficulties, lack of time, or objections from the treating specialists, workshops directed towards communicating with multiple persons and managing consultation time might help. Moreover, appointments additional to regular consultations are possible, as well as the implementation of support groups and information events focusing on the patient-caregiver unit. Extensive research on these possible interventions should be done prior to any implementation. The experiences, needs, and the role of caregivers during lung cancer treatment require further investigation.

Both patients and caregivers searched the Internet and the DLIC website broadly for additional information on lung cancer. It can therefore be concluded that there is a demand for such a service, although it was not considered a potential replacement for live consultations with the treating specialists. The positive effects on the respondents' coping and their level of satisfaction, however, show that use of such services is favorable. They should therefore be promoted as *additional* information supply sources and be part of good medical care. To prevent the use of unreliable information sources, treating specialists might refer patients and caregivers to reliable and objective websites (with online experts). James et al have already reported evidence supporting this approach as being (surprisingly) a wish of both patients and caregivers [53].

Next to referral to a specialized oncology nurse for additional information and support, the development of direct personal email contact with the nurse or with the treating specialist may also be considered, for those reluctant to use the Internet and consult online experts. There are, of course, barriers and advantages to such communication modes [54]. Barriers might be the lack of Internet access and peer pressure, as well as the absence of training or ability to use email and concerns about junk mail, privacy, and security [54]. Advantages are numerous, such as speed, efficiency, and productivity [54], and, as illustrated in our study results, satisfaction and relief of anxiety among patients and caregivers. The specialists' and nurses' perceptions and the feasibility of direct email contact should nevertheless first be investigated before future implementation.

Conclusions

Lung cancer patients and especially their caregivers use the Internet and the interactive webpage of the DLIC website because they want additional information on top of what they have received from their treating specialists. The information search also helps them to cope with lung cancer. The Internet and the DLIC's interactive page are therefore valuable complementary modes of information supply. Because the DLIC online expert is not able to answer patient-specific questions, using email contact between patients/caregivers and treating specialists or specialized oncology nurses might be considered in case of urgent questions, next to referring them to reliable sources of information.

Acknowledgments

The authors thank VA Janes for writing support. Our study was financially supported by a grant from ZonMw project number 32560015.

Conflicts of Interest

None declared.

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin 2009 Jul;59(4):225-249 [FREE Full text] [doi: 10.3322/caac.20006] [Medline: 19474385]
2. Boyle P, Autier P, Bartelink H, Baselga J, Boffetta P, Burn J, et al. European Code Against Cancer and scientific justification: third version (2003). Ann Oncol 2003 Jul;14(7):973-1005 [FREE Full text] [Medline: 12853336]
3. Nederlandse Kankerregistratie CoK. Tabellen: Prevalentie, Incidentie, Sterfte: Onderste luchtwegen; Landelijk; Man & Vrouw, Invasief. 2014. URL: http://www.cijfersoverkanker.nl/selecties/sterfte_onderste_luchtwegen/img52d00b7ad3bac [accessed 2014-01-10] [WebCite Cache ID 6MWZPnXVc]
4. Longkanker Informatiecentrum. Homepage. URL: http://www.longkanker.info/ [accessed 2010-12-21] [WebCite Cache ID 5v8jae1HR]
5. Linssen C, The AM, Lammers E, Festen J, Postmus PE. Gebruik van website over longkanker door patiënten en naasten. Nederlands Tijdschrift voor Oncologie 2006;3:102-106.
6. Linssen C, Schook RM, The AM, Lammers E, Festen J, Postmus PE. A web site on lung cancer: who are the users and what are they looking for? J Thorac Oncol 2007 Sep;2(9):813-818. [doi: 10.1097/JTO.0b013e31811f472a] [Medline: 17805058]
7. Longkanker Informatiecentrum Nederland. Statistieken 2010 website gebruik. Intern rapport. Baarn 2010:1-10.
8. Schook RM, Linssen C, Festen J, Schramel FM, Lammers E, Zaanen P. Website visitors asking questions online to lung cancer specialists: what do they want to know? Interact J Med Res 2013;2(2):e-15. [doi: 10.2196/ijmr.1749]
9. Manfredi C, Czaja R, Price J, Buis M, Janiszewski R. Cancer patients' search for information. J Natl Cancer Inst Monogr 1993;14:93-104.
10. Mills ME, Sullivan K. The importance of information giving for patients newly diagnosed with cancer: a review of the literature. J Clin Nurs 1999 Nov;8(6):631-642. [Medline: 10827609]
11. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371. [Medline: 15224975]
12. Fogel J, Albert SM, Schnabel F, Ditkoff BA, Neugut AI. Use of the Internet by women with breast cancer. J Med Internet Res 2002 Apr;4(2):E-9. [doi: 10.2196/jmir.4.2.e9]
13. Mancini J, Noguès C, Adenis C, Berthet P, Bonadona V, Chompret A, et al. Patients' characteristics and rate of Internet use to obtain cancer information. J Public Health (Oxf) 2006 Sep;28(3):235-237 [FREE Full text] [doi: 10.1093/pubmed/fdl019] [Medline: 16809788]
14. Ginossar T. Online participation: a content analysis of differences in utilization of two online cancer communities by men and women, patients and family members. Health Commun 2008;23(1):1-12. [doi: 10.1080/10410230701697100] [Medline: 18443988]
15. Jansen J, van Weert JC, Wijngaards-de Meij L, van Dulmen S, Heeren TJ, Bensing JM. The role of companions in aiding older cancer patients to recall medical information. Psychooncology 2010 Feb;19(2):170-179. [doi: 10.1002/pon.1537] [Medline: 19319855]
16. Kinnane NA, Milne DJ. The role of the Internet in supporting and informing carers of people with cancer: a literature review. Support Care Cancer 2010 Sep;18(9):1123-1136. [doi: 10.1007/s00520-010-0863-4] [Medline: 20336326]
17. Davidson JE, Powers K, Hedayat KM, Tieszen M, Kon AA, Shepard E, American College of Critical Care Medicine Task Force 2004-2005, Society of Critical Care Medicine. Clinical practice guidelines for support of the family in the patient-centered intensive care unit: American College of Critical Care Medicine Task Force 2004-2005. Crit Care Med 2007 Feb;35(2):605-622. [doi: 10.1097/01.CCM.0000254067.14607.EB] [Medline: 17205007]

18. Koninklijke Nederlandsche Maatschappij tot bevordering der Geneeskunst K. Richtlijnen inzake het omgaan met medische gegevens. KNMG 2010:1-50.
19. Maguire P. Improving communication with cancer patients. Eur J Cancer 1999 Dec;35(14):2058-2065. [Medline: 10711246]
20. Simpson M, Buckman R, Stewart M, Maguire P, Lipkin M, Novack D, et al. Doctor-patient communication: the Toronto consensus statement. BMJ 1991 Nov 30;303(6814):1385-1387 [FREE Full text] [Medline: 1760608]
21. Faulkner A. Talking to Cancer Patients and Their Relatives. Oxford: Oxford University Press; 1994.
22. Ryan PJ, Howell V, Jones J, Hardy EJ. Lung cancer, caring for the caregivers. A qualitative study of providing pro-active social support targeted to the carers of patients with lung cancer. Palliat Med 2008 Apr;22(3):233-238. [doi: 10.1177/0269216307087145] [Medline: 18477717]
23. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006;3(2):77-101.
24. VERBI GmbH. Verbi Software. 2011. Website on MAXQDA - The Art of Text Analysis - Qualitative Data Analysis Software URL: http://www.maxqda.com/contact [accessed 2014-01-10] [WebCite Cache ID 6MWZy6eLV]
25. Mays N, Pope C. Rigour and qualitative research. BMJ 1995 Jul 8;311(6997):109-112 [FREE Full text] [Medline: 7613363]
26. Johnson JD. Cancer-related information seeking. Cresskill, NJ: Hampton Press; 1997.
27. Folkman S. Positive psychological states and coping with severe stress. Soc Sci Med 1997 Oct;45(8):1207-1221. [Medline: 9381234]
28. Hilton BA. The phenomenon of uncertainty in women with breast cancer. Issues Ment Health Nurs 1988;9(3):217-238. [Medline: 3198380]
29. Mellink WA, Dulmen AM, Wiggers T, Spreeuwenberg PM, Eggermont AM, Bensing JM. Cancer patients seeking a second surgical opinion: results of a study on motives, needs, and expectations. J Clin Oncol 2003 Apr 15;21(8):1492-1497. [doi: 10.1200/JCO.2003.12.058] [Medline: 12697872]
30. Tattersall MH, Dear RF, Jansen J, Shepherd HL, Devine RJ, Horvath LG, et al. Second opinions in oncology: the experiences of patients attending the Sydney Cancer Centre. Med J Aust 2009 Aug 17;191(4):209-212. [Medline: 19705981]
31. Ong LM, de Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. Soc Sci Med 1995 Apr;40(7):903-918. [Medline: 7792630]
32. Bee PE, Barnes P, Luker KA. A systematic review of informal caregivers' needs in providing home-based end-of-life care to people with cancer. J Clin Nurs 2009 May;18(10):1379-1393. [doi: 10.1111/j.1365-2702.2008.02405.x] [Medline: 18624779]
33. Davidson R, Mills ME. Cancer patients' satisfaction with communication, information and quality of care in a UK region. Eur J Cancer Care (Engl) 2005 Mar;14(1):83-90. [doi: 10.1111/j.1365-2354.2005.00530.x] [Medline: 15698390]
34. Dolce MC. The Internet as a source of health information: experiences of cancer survivors and caregivers with healthcare providers. Oncol Nurs Forum 2011 May;38(3):353-359. [doi: 10.1188/11.ONF.353-359] [Medline: 21531685]
35. Jones R, Pearson J, McGregor S, Gilmour WH, Atkinson JM, Barrett A, et al. Cross sectional survey of patients' satisfaction with information about cancer. BMJ 1999 Nov 6;319(7219):1247-1248 [FREE Full text] [Medline: 10550091]
36. Soothill K, Morris SM, Harman JC, Francis B, Thomas C, McIllmurray MB. Informal carers of cancer patients: what are their unmet psychosocial needs? Health Soc Care Community 2001 Nov;9(6):464-475. [Medline: 11846826]
37. Soothill K, Morris SM, Thomas C, Harman JC, Francis B, McIllmurray MB. The universal, situational, and personal needs of cancer patients and their main carers. Eur J Oncol Nurs 2003 Mar;7(1):5-13; discussion 14. [Medline: 12849570]
38. Ryan PJ, Howell V, Jones J, Hardy EJ. Lung cancer, caring for the caregivers. A qualitative study of providing pro-active social support targeted to the carers of patients with lung cancer. Palliat Med 2008 Apr;22(3):233-238. [doi: 10.1177/0269216307087145] [Medline: 18477717]
39. Harrison J, Haddad P, Maguire P. The impact of cancer on key relatives: a comparison of relative and patient concerns. Eur J Cancer 1995 Oct;31A(11):1736-1740. [Medline: 8541091]
40. Burleson BR. Introduction to the special issue: Psychological Mediators of sex difference in emotional support. Communication Reports 2002;15:1-4.
41. Wright K. Social support within an online cancer community: An assessment of emotional support, perceptions of advantages and disadvantages, and motives for using the community from a communication perspective. Journal of Applied Communication 2002;30:195-209.
42. Morris SM, Thomas C. The carer's place in the cancer situation: where does the carer stand in the medical setting? Eur J Cancer Care (Engl) 2001 Jun;10(2):87-95. [Medline: 11829054]
43. Thomas C, Morris SM, Harman JC. Companions through cancer: the care given by informal carers in cancer contexts. Soc Sci Med 2002 Feb;54(4):529-544. [Medline: 11848273]
44. Helft PR, Hlubocky F, Daugherty CK. American oncologists' views of internet use by cancer patients: a mail survey of American Society of Clinical Oncology members. J Clin Oncol 2003 Mar 1;21(5):942-947. [Medline: 12610198]
45. Fried TR, Bradley EH, O'Leary JR, Byers AL. Unmet desire for caregiver-patient communication and increased caregiver burden. J Am Geriatr Soc 2005 Jan;53(1):59-65. [doi: 10.1111/j.1532-5415.2005.53011.x] [Medline: 15667377]
46. Nikoletti S, Kristjanson LJ, Tataryn D, McPhee I, Burt L. Information needs and coping styles of primary family caregivers of women following breast cancer surgery. Oncol Nurs Forum 2003;30(6):987-996. [doi: 10.1188/03.ONF.987-996] [Medline: 14603356]

47. Zhang AY, Siminoff LA. Silence and cancer: why do families and patients fail to communicate? Health Commun 2003;15(4):415-429. [doi: 10.1207/S15327027HC1504_03] [Medline: 14527866]
48. Matsuyama RK, Kuhn LA, Molisani A, Wilson-Genderson MC. Cancer patients' information needs the first nine months after diagnosis. Patient Educ Couns 2013 Jan;90(1):96-102. [doi: 10.1016/j.pec.2012.09.009]
49. Mistry A, Wilson S, Priestman T, Damery S, Haque M. How do the information needs of cancer patients differ at different stages of the cancer journey? A cross-sectional survey. JRSM Short Rep 2010;1(4):30 [FREE Full text] [doi: 10.1258/shorts.2010.010032] [Medline: 21103122]
50. Fox S, Fallows D. Pew Research Center. Health searches and email have become more commonplace, but there is room for improvement in searches and overall internet access URL: http://www.pewinternet.org/Reports/2003/Internet-Health-Resources.aspx [accessed 2010-12-21] [WebCite Cache ID 5v8iu2JuK]
51. DuBenske LL, Wen KY, Gustafson DH, Guarnaccia CA, Cleary JF, Dinauer SK, et al. Caregivers' differing needs across key experiences of the advanced cancer disease trajectory. Palliat Support Care 2008 Sep;6(3):265-272 [FREE Full text] [doi: 10.1017/S1478951508000400] [Medline: 18662420]
52. DuBenske LL, Chih MY, Gustafson DH, Dinauer S, Cleary JF. Caregivers' participation in the oncology clinic visit mediates the relationship between their information competence and their need fulfillment and clinic visit satisfaction. Patient Educ Couns 2010 Dec;81 Suppl:S94-S99 [FREE Full text] [doi: 10.1016/j.pec.2010.08.022] [Medline: 20880656]
53. James N, Daniels H, Rahman R, McConkey C, Derry J, Young A. A study of information seeking by cancer patients and their carers. Clin Oncol (R Coll Radiol) 2007 Jun;19(5):356-362. [doi: 10.1016/j.clon.2007.02.005] [Medline: 17399963]
54. Dilts D, Ridner SH, Franco A, Murphy B. Patients with cancer and e-mail: implications for clinical communication. Support Care Cancer 2009 Aug;17(8):1049-1056. [doi: 10.1007/s00520-008-0538-6] [Medline: 19050941]

Abbreviations

DLIC: Dutch Lung Cancer Information Center
NSCLC: non–small cell lung cancer
SCLC: small cell lung cancer

Edited by G Eysenbach; submitted 23.07.13; peer-reviewed by N Kinnane, A Batenburg; comments to author 19.09.13; revised version received 23.10.13; accepted 09.01.14; published 04.02.14.

Please cite as:
Schook RM, Linssen C, Schramel FMNH, Festen J, Lammers E, Smit EF, Postmus PE, Westerman MJ
Why Do Patients and Caregivers Seek Answers From the Internet and Online Lung Specialists? A Qualitative Study
J Med Internet Res 2014;16(2):e37
URL: http://www.jmir.org/2014/2/e37/
doi:10.2196/jmir.2842
PMID:24496139

Original Paper

Multiple Information Sources and Consequences of Conflicting Information About Medicine Use During Pregnancy: A Multinational Internet-Based Survey

Katri Hämeen-Anttila[1], PhD; Hedvig Nordeng[2], PhD; Esa Kokki[1], PhD; Johanna Jyrkkä[1], PhD; Angela Lupattelli[2]; Kirsti Vainio[3], PhD; Hannes Enlund[1], PhD

[1]Assessment of Pharmacotherapies, Finnish Medicines Agency, Fimea, Finland
[2]School of Pharmacy, University of Oslo, Oslo, Norway
[3]School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

Corresponding Author:
Johanna Jyrkkä, PhD
Assessment of Pharmacotherapies
Finnish Medicines Agency
PO Box
Fimea, 00034
Finland
Phone: 358 295223514
Fax: 358 295223001
Email: johanna.jyrkka@fimea.fi

Abstract

Background: A wide variety of information sources on medicines is available for pregnant women. When using multiple information sources, there is the risk that information will vary or even conflict.

Objective: The objective of this multinational study was to analyze the extent to which pregnant women use multiple information sources and the consequences of conflicting information, and to investigate which maternal sociodemographic, lifestyle, and medical factors were associated with these objectives.

Methods: An anonymous Internet-based questionnaire was made accessible during a period of 2 months, on 1 to 4 Internet websites used by pregnant women in 5 regions (Eastern Europe, Western Europe, Northern Europe, Americas, Australia). A total of 7092 responses were obtained (n=5090 pregnant women; n=2002 women with a child younger than 25 weeks). Descriptive statistics and logistic regression analysis were used.

Results: Of the respondents who stated that they needed information, 16.16% (655/4054) used one information source and 83.69% (3393/4054) used multiple information sources. Of respondents who used more than one information source, 22.62% (759/3355) stated that the information was conflicted. According to multivariate logistic regression analysis, factors significantly associated with experiencing conflict in medicine information included being a mother (OR 1.32, 95% CI 1.11-1.58), having university (OR 1.33, 95% CI 1.09-1.63) or other education (OR 1.49, 95% CI 1.09-2.03), residing in Eastern Europe (OR 1.52, 95% CI 1.22-1.89) or Australia (OR 2.28, 95% CI 1.42-3.67), use of 3 (OR 1.29, 95% CI 1.04-1.60) or >4 information sources (OR 1.82, 95% CI 1.49-2.23), and having ≥2 chronic diseases (OR 1.49, 95% CI 1.18-1.89). Because of conflicting information, 43.61% (331/759) decided not to use medication during pregnancy, 30.30% (230/759) sought a new information source, 32.67% (248/759) chose to rely on one source and ignore the conflicting one, 25.03% (190/759) became anxious, and 2.64% (20/759) did nothing. Factors significantly associated with not using medication as a consequence of conflicting information were being pregnant (OR 1.75, 95% CI 1.28-2.41) or experiencing 3-4 health disorders (OR 1.99, 95% CI 1.10-3.58). Women with no chronic diseases were more likely not to take medicines than women with ≥2 chronic diseases (OR 2.22, 95% CI 1.47-3.45). Factors significantly associated with becoming anxious were >4 information sources (OR 2.67, 95% CI 1.70-4.18) and residing in Eastern Europe (OR 0.57, 95% CI 0.36-0.90).

Conclusions: Almost all the pregnant women used multiple information sources when seeking information on taking medicines during pregnancy and one-fifth obtained conflicting information, leading to anxiety and the decision not to use the medication. Regional, educational, and chronic disease characteristics were associated with experiencing conflicting information and influenced the decision not to use medication or increased anxiety. Accurate and uniform teratology information should be made more available to the public.

(J Med Internet Res 2014;16(2):e60) doi:10.2196/jmir.2939

KEYWORDS

pharmaceutical preparations; pregnancy; access to information; information seeking behavior; information dissemination; questionnaires; Internet; international

Introduction

Use of both prescription medicines and over-the-counter medicines during pregnancy is common [1-3]. Pregnant women tend to be cautious about using medication and tend to have unrealistic perceptions of drug-related teratogenic risks [4,5]. This may result in nonadherence to therapy or even to the unjustified termination of a pregnancy [6,7]. Relatively few studies focus on adherence to drug therapy during pregnancy [8]. Those that do have shown that the overall adherence rate during pregnancy is approximately 40% [9-11]. Adherence to chronic medication seems higher than to short-term medication or symptomatic treatment [10], but the opposite has also been reported in treatment of asthma [12] and depression [13]. Poor adherence among pregnant women is typically seen as reduction of the dose, total discontinuation of medication [9,14,15], or forgetting to take the medication [16,17]. Conflicting information from different sources may add to the uncertainty about whether or not to use medication [5,13].

Pregnant women use health care professionals, most commonly physicians, but also pharmacists and nurses, as primary information sources [5,18-20]. The Internet is also a widely used information source [5,18,19]. A recent multinational study showed that 70% of the responding pregnant women who indicated a need for information used the Internet as an information source, varying from 44% in Canada to 90% in Russia [19]. Other information sources include patient information leaflets, family and friends, drug information centers, books, and magazines.

This indicates the existence of a wide variety of formal and informal information sources, which pregnant women may consult simultaneously. In fact, Henry and Crowther [20] reported in their review that 4 information sources are used on average, with one-quarter of respondents consulting more than 5 sources (range 0-11 information sources among different studies). With an increasing number of information sources, there is also an increased risk that information will vary or even conflict. In a Norwegian study of pregnant women who used several information sources, 25% reported the presence of conflicting information between sources [5]. Thus, it may be hypothesized that conflicting information increases the possibility that a medicine will not be used during pregnancy even if it is safe and important to the pregnant woman and the unborn child.

This multinational survey aims to identify the extent to which pregnant women use multiple information sources and the consequences of the presence of conflicting information. Furthermore, it aims to investigate which maternal sociodemographic, lifestyle, and medical factors are associated with these objectives.

Methods

Study Design

This study forms part of a multinational Internet-based survey on medication use in pregnancy, which investigates medicine use, health disorders, and chronic diseases during pregnancy; perceptions of risks and attitudes toward using medicines; and needs for information [11]. The survey was conducted in 5 regions: Eastern Europe (Croatia, Poland, Russia, Serbia, Slovenia), Western Europe (Austria, France, Italy, the Netherlands, Switzerland, United Kingdom), Northern Europe (Finland, Iceland, Norway, Sweden), Americas (Canada, United States, South America), and Australia.

An anonymous self-completed Internet-based questionnaire [21] was posted on 1 to 4 Internet websites used by pregnant women in different countries. Originally developed at the University of Oslo [5], the questionnaire was translated into the respective languages of the participating countries. The questionnaire was piloted in 4 countries (pilot responses are not included in the study). Only minor modifications were made. During the pilot, the usability and technical functionality of the electronic questionnaire were also tested. Adaptive questioning was used to reduce the number and complexity of the questions. The questionnaire included various topics concerning medicine use and health during pregnancy, as well as attitudes toward using medicines in general and during pregnancy. In this study, data were used that had been gathered based on responses to questions concerning the need for information and reported information sources.

The questionnaire was accessible during a period of 2 months in each country, from October 1, 2011 to February 29, 2012. Pregnant women and women with a child younger than 25 weeks old were eligible to participate in this study. Respondents were advised to answer questions related to their current or latest pregnancy. The participants were also asked to read the study description, along with the study objectives and other relevant information, before being given access to the online questionnaire. Reading the study description and confirming a wish to participate were considered the equivalent of giving informed consent. Thereafter, the woman was given access to the online questionnaire. No personal identifiable information was collected. Ethical approval was sought and granted from the Norwegian Regional Ethics Committee.

In each country, the study population was compared to the birthing population by using national or population-based statistics to evaluate external validity [19]. Overall, the mean age of the study populations in each country was close to the mean age of the target populations. Respondents were somewhat better educated than average women, and the percentage of primiparity was higher among respondents than among most national populations. Of the respondents, 57.16% (4054/7092) reported a need for information on medicines during their pregnancy [19]. The most commonly used information sources were health care professionals (physicians: 73.14%, 2965/4054; pharmacy personnel: 46.18%, 1872/4054; midwifes or nurses: 32.51%, 1318/4054) and the Internet (59.50%, 2412/4054). A detailed description of information needs and the information sources used is given elsewhere [19].

Main Outcome Measures

A list of commonly used sources was given to explore the number and type of information sources used. Respondents were also given the opportunity to mention additional sources. The basis for this list was taken from a previous study by the authors [5] and adjusted in relation to the information sources available in the participating countries. These were further categorized into formal information sources (including physicians, pharmacists, midwifes, nurses, drug handbooks, information leaflets, and drug information centers) and informal information sources (including Internet, family and friends, magazines, media, books, and herbal shop personnel). Drug information centers refer to medicine information services (also known drug information call centers) where people can call or otherwise contact health care professionals and inquire about medicines.

The need for information was assessed with the question, “Did you need information on medicines during the course of your pregnancy?” Respondents who indicated a need for information (n=4054) were further asked if the information they had found from various sources was uniform. The respondents could choose from 4 responses: (1) yes, completely similar; (2) yes, on the whole (only the wording or level of detail was somewhat different); (3) no, part of the information was different; and (4) no, the information was completely contradictory. These answers were further classified into the categories “yes” (2 former categories) and “no” (2 latter categories).

If the respondent’s answers fell into any of the “no” categories, she was further asked about her subsequent actions with the question: “If there were discrepancies among the sources, what did this mean to you? (You may tick more than one answer).” The 5 possible responses were: (1) nothing, (2) I became anxious, (3) I decided not to use the medication, (4) I sought a new information source, and (5) I chose to rely on one source only, and to ignore the conflicting one.

Background Variables

The following background variables were used: the status of the women (pregnant or had given birth at the time of the study), age, parity, marital status, educational level, region of residence, number of chronic diseases, experienced health disorders, and number of information sources used.

Statistical Analysis

We used SPSS version 20 (IBM Corp, Armonk, NY, USA) for statistical analyses. Cross-tabulation and Pearson’s chi-square test were used as univariate analysis when analyzing categorical variables. A *P* value of <.05 was considered to be statistically significant. For multivariate analysis, logistic regression analysis including odds ratios (OR) and 95% confidence intervals (95% CIs) were used when measuring the association of maternal sociodemographics and lifestyle characteristics with the experience of conflicting information and with the consequences of the presence of conflicting information during pregnancy. The stepwise method (forward conditional) was used to select the variables in the final model. Multivariate analyses were conducted for the consequences “did not use medication” and “became anxious.” The Hosmer and Lemeshow test [22] was used to assess goodness-of-fit of the final multivariate models and the models were found robust (*P* values >.05).

Results

Summary

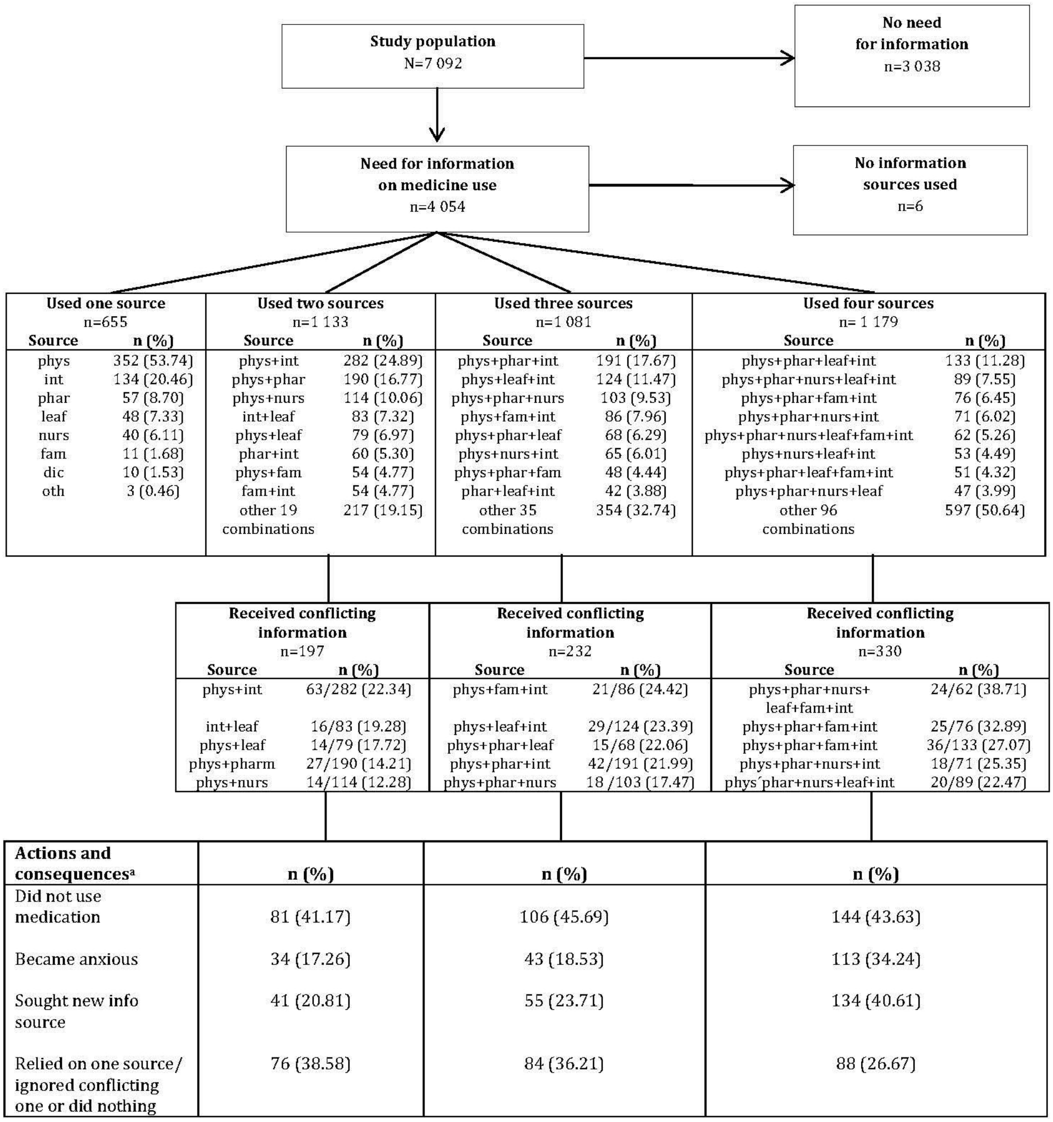
A total of 9615 women in various countries accessed the online open survey. Of these women, 9483 (98.63%) agreed to participate and filled in the questionnaire. Of the related responses, 7092 of 9483 (74.79%) were eligible for this study, including 5090 (53.67%) responses from pregnant women and 2002 (21.11%) from women with a child younger than 25 weeks.

Information Sources on Medicines

On average, the respondents used 3 different information sources on medicines (range 0-8). Of these, 16.16% (655/4054) used 1 source, 27.95% (1133/4054) used 2 sources, 26.66% (1081/4054) used 3 sources, and 29.08% (1179/4054) used 4 or more different information sources (Figure 1). A physician was used as the sole information source by 53.74% (352/655) of the women who used 1 information source, followed by the Internet (20.46%, 134/655) and a pharmacist (8.70%, 57/655) (Figure 1). Physicians (77.01%, 3613/3393), the Internet (67.14%, 2278/3393), and pharmacists (53.49%, 1815/3393) also featured most often among the combinations of various information sources. Other sources included nurses or midwives; patient information leaflets; drug information centers; family or friends; books, magazines, or media; and herbal shops.

Of the women needing information, 94.44% (3829/4054) used formal information sources (including physicians, pharmacists, midwifes, nurses, drug handbooks, patient information leaflets, and drug information centers) and 67.04% (2718/4054) used informal information sources (including the Internet, family and friends, magazines, media, books, and herbal shop personnel). The Internet was used by 59.50% (2412/4054) of women needing information. Both formal and informal information sources were used by 61.64% (2499/4054).

Figure 1. Need for information on medicine use including actions and consequences ensuing from conflicting information. Phys: physicians; phar: pharmacists; nurs: nurses or midwives; leaf: patient information leaflets; dic: drug information centers; int: Internet; fam: family or friends.



Used one source n=655 Source	n (%)
phys	352 (53.74)
int	134 (20.46)
phar	57 (8.70)
leaf	48 (7.33)
nurs	40 (6.11)
fam	11 (1.68)
dic	10 (1.53)
oth	3 (0.46)

Used two sources n=1 133 Source	n (%)
phys+int	282 (24.89)
phys+phar	190 (16.77)
phys+nurs	114 (10.06)
int+leaf	83 (7.32)
phys+leaf	79 (6.97)
phar+int	60 (5.30)
phys+fam	54 (4.77)
fam+int	54 (4.77)
other 19 combinations	217 (19.15)

Used three sources n=1 081 Source	n (%)
phys+phar+int	191 (17.67)
phys+leaf+int	124 (11.47)
phys+phar+nurs	103 (9.53)
phys+fam+int	86 (7.96)
phys+phar+leaf	68 (6.29)
phys+nurs+int	65 (6.01)
phys+phar+fam	48 (4.44)
phar+leaf+int	42 (3.88)
other 35 combinations	354 (32.74)

Used four sources n= 1 179 Source	n (%)
phys+phar+leaf+int	133 (11.28)
phys+phar+nurs+leaf+int	89 (7.55)
phys+phar+fam+int	76 (6.45)
phys+phar+nurs+int	71 (6.02)
phys+phar+nurs+leaf+fam+int	62 (5.26)
phys+nurs+leaf+int	53 (4.49)
phys+phar+leaf+fam+int	51 (4.32)
phys+phar+nurs+leaf	47 (3.99)
other 96 combinations	597 (50.64)

Received conflicting information n=197 Source	n (%)
phys+int	63/282 (22.34)
int+leaf	16/83 (19.28)
phys+leaf	14/79 (17.72)
phys+pharm	27/190 (14.21)
phys+nurs	14/114 (12.28)

Received conflicting information n=232 Source	n (%)
phys+fam+int	21/86 (24.42)
phys+leaf+int	29/124 (23.39)
phys+phar+leaf	15/68 (22.06)
phys+phar+int	42/191 (21.99)
phys+phar+nurs	18 /103 (17.47)

Received conflicting information n=330 Source	n (%)
phys+phar+nurs+leaf+fam+int	24/62 (38.71)
phys+phar+fam+int	25/76 (32.89)
phys+phar+fam+int	36/133 (27.07)
phys+phar+nurs+int	18/71 (25.35)
phys´phar+nurs+leaf+int	20/89 (22.47)

Actions and consequences[a]	n (%)	n (%)	n (%)
Did not use medication	81 (41.17)	106 (45.69)	144 (43.63)
Became anxious	34 (17.26)	43 (18.53)	113 (34.24)
Sought new info source	41 (20.81)	55 (23.71)	134 (40.61)
Relied on one source/ ignored conflicting one or did nothing	76 (38.58)	84 (36.21)	88 (26.67)

Conflicting Information From Different Sources

Of the respondents who used more than 1 information source, 22.62% (759/3355) indicated that information was conflicting. The proportion of women receiving conflicting information increased along with the number of sources. This varied from 2 information sources, of which 17.39% (197/1133) consisted of conflicting information, to 4 or more information sources, of which 27.99% (330/1179) contained conflicting information (Figure 1).

The experience of conflicting information was greatest among the mothers who had university education or other education (Table 1). Furthermore, high numbers of information sources and high numbers of chronic diseases were associated with the experienced information conflicts. There were also regional differences: women from Eastern Europe and Australia reporting experiencing conflicting information more often than women in other regions.

Table 1. Univariate and multivariate analysis of factors associated with conflict of information (n=3355).

Variable	n	Information type, n (%)		Contradictory information			
		Similar information	Contradictory information	Univariate logistic regression		Multivariate logistic regression[a]	
				OR (95% CI)	*P*	OR (95% CI)	*P*
Status of the woman					.001		.002
Pregnant	2379	1878 (78.9)	501 (21.1)	1.00		1.00	
Mother	976	718 (73.6)	258 (26.4)	1.35 (1.13-1.60)		1.32 (1.11-1.58)	
Age (years)					.08	—	
≤24	615	496 (80.7)	119 (19.3)	0.77 (0.62-0.97)			
25-34	2173	1659 (76.3)	514 (23.7)	1.00			
≥35	531	412 (77.6)	119 (22.4)	0.93 (0.74-1.17)			
Parity					.21	—	
Primiparous	1928	1507 (78.2)	421 (21.8)	1.00			
≥1 previous children	1427	1089 (76.3)	338 (23.7)	1.11 (0.94-1.31)			
Marital status					.15	—	
Married or cohabitant	3149	2445 (77.6)	704 (22.4)	1.00			
Other	206	151 (73.3)	55 (26.7)	1.27 (0.92-1.74)			
Education level					.003		.02
Primary school	153	124 (81.0)	29 (19.0)	1.03 (0.66-1.59)		1.02 (0.65-1.62)	
High/secondary school	884	720 (81.4)	164 (18.6)	1.00		1.00	
University	1990	1507 (75.7)	483 (24.3)	1.41 (1.15-1.72)		1.33 (1.09-1.63)	
Other Education	328	245 (74.7)	83 (25.3)	1.49 (1.10-2.01)		1.49 (1.09-2.03)	
Region of residence					<.001		<.001
Western Europe	1030	838 (81.4)	192 (18.6)	1.00		1.00	
Eastern Europe	919	676 (73.6)	243 (26.4)	1.57 (1.27-1.95)		1.52 (1.22-1.89)	
Northern Europe	1046	817 (78.1)	229 (21.9)	1.22 (0.99-1.52)		1.15 (0.92-1.44)	
Americas	274	211 (77.0)	63 (23.0)	1.30 (0.94-1.80)		1.28 (0.92-1.77)	
Australia	86	54 (62.8)	32 (37.2)	2.59 (1.63-4.11)		2.28 (1.42-3.67)	
Experienced health disorders					.63	—	
0-2	310	246 (79.4)	64 (20.6)	1.00			
3-4	1003	770 (76.8)	233 (23.2)	0.90 (0.67-1.20)			
≥5	2028	1572 (77.5)	456 (22.5)	1.04 (0.87-1.25)			
Number of chronic diseases					<.001		.002
0	2294	1816 (79.2)	478 (20.8)	1.00		1.00	
1	592	446 (75.3)	146 (24.7)	1.24 (1.01-1.54)		1.20 (0.96-1.49)	
≥2	469	334 (71.2)	135 (28.8)	1.54 (1.23-1.92)		1.49 (1.18-1.89)	
Multiple information sources					<.001		<.001
2 sources	1112	915 (82.3)	197 (17.7)	1.00		1.00	
3 sources	1071	839 (78.3)	232 (21.7)	1.28 (1.04-1.59)		1.29 (1.04-1.60)	
≥4 sources	1172	842 (71.8)	330 (28.2)	1.82 (1.49-2.22)		1.82 (1.49-2.23)	

[a] Adjusted with the variables shown (n=3305).

Responses to Conflicting Information on Medicines

Almost half (43.6%, 332/759) of women decided not to use medication because of conflicting information; 30.3% (230/759) sought a new information source; 32.7% (248/759) chose to rely on one source and ignore the conflicting one; 25.0% (190/759) became anxious; and 2.6% (20/759) did nothing (Table 2).

According to the multivariate analysis, pregnant women (47.9%, 240/501) more often than mothers (35.3%, 91/258) reported not using medication as a consequence of conflicting information when adjusted for other variables (OR 1.75, 95% CI 1.28-2.41). This was also the case for women who had experienced 3-4 health disorders (50.2%, 117/233) compared to women with experience of 0-2 health disorders (35.9%, 23/64; OR 1.99, 95% CI 1.10-3.58). Moreover, women with no chronic diseases (48.6%, 228/478) were more likely to state that they did not use medicines than women with 2 or more chronic diseases (28.9%, 39/135; OR 2.22, 95% CI 1.47-3.45). No other significant associations were found.

Multivariate analysis also showed that women who used 4 or more information sources (34.2%, 113/330) reported becoming anxious as a consequence of conflicting information more often than women who used 2 information sources (17.2%, 34/197; OR 2.67, 95% CI 1.70-4.18) after adjustment for other variables. Furthermore, women from Eastern Europe (19.7%, 48/243) were less likely to report becoming anxious than women from Western Europe (28.1%, 54/192; OR 0.57, 95% CI 0.36-0.90). Again, no other significant associations were found.

Table 2. Consequences of discrepancies between different information sources by background variables (n=759).

Variable	n	Consequence,[a] n (%)			
		Did not use medication	Became anxious	Sought a new information source	Relied on one source and ignored conflicting one, or did nothing
Status of the woman					
Pregnant	501	240 (47.9)	118 (23.6)	135 (26.9)	161 (32.1)
Mother	258	91 (35.3)	72 (27.9)	95 (36.8)	87 (33.7)
Age (years)					
≤24	119	63 (52.9)	28 (23.5)	36 (30.3)	30 (25.2)
25-34	514	217 (42.2)	132 (25.7)	163 (31.7)	173 (33.7)
≥35	119	49 (41.2)	27 (22.7)	31 (26.0)	41 (34.5)
Parity					
Primiparous	421	185 (43.9)	105 (24.9)	129 (30.6)	135 (32.1)
≥1 previous children	338	146 (43.2)	85 (25.1)	101 (29.9)	113 (33.4)
Marital status					
Married or cohabitant	704	308 (43.7)	170 (24.1)	215 (30.6)	232 (32.9)
Other	55	23 (41.8)	20 (36.4)	15 (27.3)	16 (29.1)
Education level					
Primary school	29	11 (37.9)	5 (17.2)	7 (24.1)	14 (48.3)
High/secondary school	164	81 (49.4)	46 (28.0)	27 (16.5)	54 (32.9)
University	483	206 (42.6)	110 (22.8)	168 (34.8)	157 (32.5)
Other education	83	33 (39.7)	29 (34.9)	28 (33.7)	21 (25.3)
Region of residence					
Western Europe	192	84 (43.7)	54 (28.1)	43 (22.4)	61 (31.8)
Eastern Europe	243	118 (48.6)	48 (19.7)	87 (35.8)	67 (27.6)
Northern Europe	229	95 (41.5)	58 (25.3)	65 (28.4)	91 (39.7)
Americas	63	25 (39.7)	21 (33.3)	21 (33.3)	19 (30.2)
Australia	32	9 (28.1)	9 (28.1)	14 (43.7)	10 (31.2)
Experienced health disorders					
0-2	64	23 (35.9)	11 (17.2)	20 (31.2)	19 (29.7)
3-4	233	117 (50.2)	50 (21.4)	20 (8.6)	75 (32.2)
≥5	456	189 (41.4)	128 (28.1)	137 (30.0)	153 (33.6)
Number of chronic diseases					
0	478	228 (47.7)	114 (23.8)	122 (25.5)	147 (30.7)
1	146	64 (43.8)	36 (24.6)	56 (38.3)	44 (30.1)
≥2	135	39 (28.9)	40 (29.6)	52 (38.5)	57 (42.2)
Multiple information sources					
2	197	81 (41.1)	34 (17.2)	41 (20.8)	76 (38.6)
3	232	106 (45.7)	43 (18.5)	55 (23.7)	84 (36.2)
≥4	330	144 (43.6)	113 (34.2)	134 (40.6)	88 (26.7)

[a] Respondents could choose more than 1 answer.

Discussion

Principal Findings

According to this study, most (83.69%, 3393/4054) pregnant women reported used multiple information sources when seeking information on medicine use during pregnancy. Of these, approximately one-fifth (18.7%, 759/4054) reported that information received from multiple sources was conflicting. Experiencing conflicting information was associated with being a mother (compared to being pregnant), having university or other education, having chronic diseases, and using multiple information sources. Furthermore, regional differences were found. As a consequence, substantial numbers of women who received conflicting information decided not to use their medication. Increased anxiety was also common. These results confirm earlier results recorded in studies of adherence problems [8,12-15].

According to our results and the results of previous studies, pregnant women actively seek information on medicines and health from various information sources [5,18-20]. To our knowledge, our study is the first to reveal the extent and wide variety of combinations of information sources used by pregnant women from 5 regions of the world involved in this study. However, in each case, we do not know which source came first, in which order the information was sought, or the actual content of the information received. There is evidence of a lack of consistency between information on drug safety during pregnancy based on different information sources [23-25]. For example, according to the study by Frost Widnes and Schjøtt [24], the Norwegian Compendium on Product Monographs gives advice which is more restrictive than that given by drug information centers. Furthermore, patient information leaflets have been shown to include varying information concerning medicine use during pregnancy when different brand names of the same active substance are compared [23]. Thus, it is not surprising that we found that the number of information sources was associated with the experience of conflict in medical information.

Importantly, pregnant women may find safe lists for medications that can be taken during pregnancy from the Internet that have no basis in evidence [25]. This suggests a clear risk of unnecessary use of medicines, misunderstandings, and groundless anxiety that could be avoided by the deployment of relevant, uniform, and accurate information. It has been argued that teratology information services (TIS) are effective in teratology information knowledge transfer by using evidence-based information expressed in lay language [26-28]. These TIS exist all over North America, in Australia, and many European countries.

The Internet is commonly used as a supplementary information source, before or after contacting a health care professional [29-31]. Women not only seek factual knowledge from the Internet, but also look for emotional support and encouragement, especially from other women in the same life situation [29,32]. Women determine which websites they can trust based on methods that include repetition of the same facts on various sites [29-31]. This leads to confusion about what information to trust when it proves to be inconsistent among different sites. Despite this, most women do not consult health care professionals about the health information they have retrieved [30,31]. Every health care professional, including physicians, nurses, midwifes, and pharmacists, should be active in asking pregnant women whether or not they have sought information on the Internet and about where and what kind of information they have found and if the information found has raised any questions. In addition, the basic education of all health care professionals should include information on reliable online information sources, targeted to medicine users as well as professionals, and about how such information sources should be discussed with the medicine user.

Limitations

This study has several limitations that should be considered when interpreting the results. Because the questionnaire was only available through the Internet, we are unable to calculate a conventional response rate. This made a comparison with the birthing population of each of the participating countries necessary to determine representativeness. However, such comparisons could not be made for South American countries because birthing population reports were not available. Furthermore, based on previous epidemiological studies using Web-based recruitment methods, a reasonable level of validity can be expected [33-35]. It should also be noted that we do not know precisely which websites the respondents used and because of the cross-sectional study design, neither do we know which information source was consulted first (eg, whether the Internet was used as the first point of enquiry or whether the women searched it for a second opinion after visiting a health care professional). Furthermore, education was categorized in the questionnaire including a possibility to choose “other education.” This is why we do not know what kind of education this includes.

The study population is fairly representative of the target populations of each participating country [19]. However, respondents were somewhat better educated and the study population had more primiparous women than the target population. Better-educated women tend to seek information and use multiple information sources [36,37], which may result in overestimating the need for information and use of multiple sources. It might also be assumed that women with medical problems or using medicines during pregnancy were more likely than others to respond to the survey.

Conclusions

Almost all the pregnant women who needed information used multiple information sources when seeking information on medicines during pregnancy. One-fifth of the respondents had obtained conflicting information on medicine use during pregnancy. Being a mother, having university education, the number of chronic diseases, and number of information sources consulted was associated with the experience of conflicting information. Furthermore, region of residence in Eastern Europe or Australia was also associated with this experience.

Conflicting information often lead to anxiety and the decision not to use medication; pregnant women with 3-4 experienced health disorders but no chronic diseases reported not using medication as a consequence of conflicting information more often than others.

Action needs to be taken to make accurate and uniform teratology information more available to the public, such as ensuring that basic education of health care professionals includes medicine use during pregnancy. It is also important to increase the availability of reliable information (eg, TIS websites). Health care professionals should actively ask pregnant women about the information they have found from various sources and discuss this information with them.

Acknowledgments

We would like to thank all the women around the world who answered the survey, as well as the national coordinators (MJ Twigg, K Zagorodnikova, AC Mårdby, ME Moretti, M Drozd, A Panchaud, A Rieutord, R Gjergja Juraski, M Odalovic, D Kennedy, G Rudolf, H Juch, JLM Passier, and I Björnsdóttir) and all website providers who contributed to the recruitment phase. Furthermore, we would like to acknowledge the Norwegian Research Council as the provider of funding for the study (grant no: 216771/F11).

Conflicts of Interest

None declared.

References

1. Werler MM, Mitchell AA, Hernandez-Diaz S, Honein MA. Use of over-the-counter medications during pregnancy. Am J Obstet Gynecol 2005 Sep;193(3 Pt 1):771-777. [doi: 10.1016/j.ajog.2005.02.100] [Medline: 16150273]
2. Daw JR, Hanley GE, Greyson DL, Morgan SG. Prescription drug use during pregnancy in developed countries: a systematic review. Pharmacoepidemiol Drug Saf 2011 Sep;20(9):895-902 [FREE Full text] [doi: 10.1002/pds.2184] [Medline: 21774029]
3. Mitchell AA, Gilboa SM, Werler MM, Kelley KE, Louik C, Hernández-Díaz S, National Birth Defects Prevention Study. Medication use during pregnancy, with particular focus on prescription drugs: 1976-2008. Am J Obstet Gynecol 2011 Jul;205(1):51.e1-51.e8 [FREE Full text] [doi: 10.1016/j.ajog.2011.02.029] [Medline: 21514558]
4. Nordeng H, Koren G, Einarson A. Pregnant women's beliefs about medications--a study among 866 Norwegian women. Ann Pharmacother 2010 Sep;44(9):1478-1484. [doi: 10.1345/aph.1P231] [Medline: 20736425]
5. Nordeng H, Ystrøm E, Einarson A. Perception of risk regarding the use of medications and other exposures during pregnancy. Eur J Clin Pharmacol 2010 Feb;66(2):207-214. [doi: 10.1007/s00228-009-0744-2] [Medline: 19841915]
6. Koren G, Pastuszak A. Prevention of unnecessary pregnancy terminations by counselling women on drug, chemical, and radiation exposure during the first trimester. Teratology 1990 Jun;41(6):657-661. [doi: 10.1002/tera.1420410602] [Medline: 2353314]
7. Lupattelli A, Spigset O, Nordeng H. Adherence to medication for chronic disorders during pregnancy: results from a multinational study. Int J Clin Pharm 2014 Feb;36(1):145-153. [doi: 10.1007/s11096-013-9864-y] [Medline: 24162929]
8. Matsui D. Adherence with drug therapy in pregnancy. Obstet Gynecol Int 2012 Dec;2012:796590 [FREE Full text] [doi: 10.1155/2012/796590] [Medline: 22242026]
9. van Trigt AM, Waardenburg CM, Haaijer-Ruskamp FM, de Jong-van den Berg LT. Questions about drugs: how do pregnant women solve them? Pharm World Sci 1994 Dec 2;16(6):254-259. [Medline: 7889024]
10. Olesen C, Søndergaard C, Thrane N, Nielsen GL, de Jong-van den Berg L, Olsen J, EuroMAP Group. Do pregnant women report use of dispensed medications? Epidemiology 2001 Sep;12(5):497-501. [Medline: 11505166]
11. Lupattelli A, Spigset S, Twigg M, Zagorodnikova K, Mårdby AC, Moretti M, et al. Medication use in pregnancy: a cross-sectional, multinational web-based study. BMJ Open 2014 (forthcoming).
12. Sawicki E, Stewart K, Wong S, Paul E, Leung L, George J. Management of asthma by pregnant women attending an Australian maternity hospital. Aust N Z J Obstet Gynaecol 2012 Apr;52(2):183-188. [doi: 10.1111/j.1479-828X.2011.01385.x] [Medline: 22141407]
13. Bonari L, Koren G, Einarson TR, Jasper JD, Taddio A, Einarson A. Use of antidepressants by pregnant women: evaluation of perception of risk, efficacy of evidence based counseling and determinants of decision making. Arch Womens Ment Health 2005 Nov;8(4):214-220. [doi: 10.1007/s00737-005-0094-8] [Medline: 15959622]
14. Einarson A, Selby P, Koren G. Abrupt discontinuation of psychotropic drugs during pregnancy: fear of teratogenic risk and impact of counselling. J Psychiatry Neurosci 2001 Jan;26(1):44-48 [FREE Full text] [Medline: 11212593]
15. Mountifield RE, Prosser R, Bampton P, Muller K, Andrews JM. Pregnancy and IBD treatment: this challenging interplay from a patients' perspective. J Crohns Colitis 2010 Jun;4(2):176-182. [doi: 10.1016/j.crohns.2009.10.001] [Medline: 21122502]
16. Sawicki E, Stewart K, Wong S, Leung L, Paul E, George J. Medication use for chronic health conditions by pregnant women attending an Australian maternity hospital. Aust N Z J Obstet Gynaecol 2011 Aug;51(4):333-338. [doi: 10.1111/j.1479-828X.2011.01312.x] [Medline: 21806573]

17. Barbour RS, Macleod M, Mires G, Anderson AS. Uptake of folic acid supplements before and during pregnancy: focus group analysis of women's views and experiences. J Hum Nutr Diet 2012 Apr;25(2):140-147. [doi: 10.1111/j.1365-277X.2011.01216.x] [Medline: 22126238]
18. Bakhireva LN, Young BN, Dalen J, Phelan ST, Rayburn WF. Patient utilization of information sources about safety of medications during pregnancy. J Reprod Med 2011 Aug;56(7-8):339-343. [Medline: 21838165]
19. Hämeen-Anttila K, Jyrkkä J, Enlund H, Nordeng H, Lupattelli A, Kokki E. Medicines information needs during pregnancy: a multinational comparison. BMJ Open 2013;3(4):e002594 [FREE Full text] [doi: 10.1136/bmjopen-2013-002594] [Medline: 23624989]
20. Henry A, Crowther C. Sources of advice on medication use in pregnancy and reasons for medication uptake and cessation during pregnancy. Aust N Z J Obstet Gynaecol 2000 May;40(2):173-175. [Medline: 10925904]
21. QuestBack program. URL: http://www.questback.com/ [accessed 2014-02-03] [WebCite Cache ID 6N6y4d78Y]
22. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd edition. New York: Wiley-Interscience; 2000.
23. Bjerrum L, Foged A. Patient information leaflets--helpful guidance or a source of confusion? Pharmacoepidemiol Drug Saf 2003 Feb;12(1):55-59. [doi: 10.1002/pds.795] [Medline: 12616848]
24. Frost Widnes SK, Schjøtt J. Advice on drug safety in pregnancy: are there differences between commonly used sources of information? Drug Saf 2008;31(9):799-806. [Medline: 18707194]
25. Peters SL, Lind JN, Humphrey JR, Friedman JM, Honein MA, Tassinari MS, et al. Safe lists for medications in pregnancy: inadequate evidence base and inconsistent guidance from Web-based information, 2011. Pharmacoepidemiol Drug Saf 2013 Mar;22(3):324-328. [doi: 10.1002/pds.3410] [Medline: 23359404]
26. Hancock RL, Koren G, Einarson A, Ungar WJ. The effectiveness of Teratology Information Services (TIS). Reprod Toxicol 2007 Feb;23(2):125-132. [doi: 10.1016/j.reprotox.2006.11.005] [Medline: 17184969]
27. Shahin I, Einarson A. Knowledge transfer and translation: examining how teratogen information is disseminated. Birth Defects Res A Clin Mol Teratol 2011 Nov;91(11):956-961. [doi: 10.1002/bdra.22851] [Medline: 21948595]
28. Frost Widnes SK, Schjøtt J. Drug use in pregnancy--physicians' evaluation of quality and clinical impact of drug information centres. Eur J Clin Pharmacol 2009 Mar;65(3):303-308. [doi: 10.1007/s00228-008-0593-4] [Medline: 19052735]
29. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 1;6(1):e7 [FREE Full text] [doi: 10.2196/jmir.6.1.e7] [Medline: 15111273]
30. Larsson M. A descriptive study of the use of the Internet by women seeking pregnancy-related information. Midwifery 2009 Feb;25(1):14-20. [doi: 10.1016/j.midw.2007.01.010] [Medline: 17408822]
31. Gao LL, Larsson M, Luo SY. Internet use by Chinese women seeking pregnancy-related information. Midwifery 2013 Jul;29(7):730-735. [doi: 10.1016/j.midw.2012.07.003] [Medline: 22958935]
32. Benn C, Budge RC, White GE. Women planning and experiencing pregnancy and childbirth: information needs and sources. Nurs Prax N Z 1999 Nov;14(3):4-15. [Medline: 11221319]
33. Ekman A, Dickman PW, Klint A, Weiderpass E, Litton JE. Feasibility of using web-based questionnaires in large population-based epidemiological studies. Eur J Epidemiol 2006;21(2):103-111. [doi: 10.1007/s10654-005-6030-4] [Medline: 16518678]
34. Huybrechts KF, Mikkelsen EM, Christensen T, Riis AH, Hatch EE, Wise LA, et al. A successful implementation of e-epidemiology: the Danish pregnancy planning study 'Snart-Gravid'. Eur J Epidemiol 2010 May;25(5):297-304 [FREE Full text] [doi: 10.1007/s10654-010-9431-y] [Medline: 20148289]
35. van Gelder MM, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol 2010 Dec 1;172(11):1292-1298 [FREE Full text] [doi: 10.1093/aje/kwq291] [Medline: 20880962]
36. Plantin L, Daneback K. Parenthood, information and support on the internet. A literature review of research on parents and professionals online. BMC Fam Pract 2009;10:34 [FREE Full text] [doi: 10.1186/1471-2296-10-34] [Medline: 19450251]
37. Khoo K, Bolt P, Babl FE, Jury S, Goldman RD. Health information seeking by parents in the Internet age. J Paediatr Child Health 2008;44(7-8):419-423. [doi: 10.1111/j.1440-1754.2008.01322.x] [Medline: 18564080]

Abbreviations

TIS: teratology information services

Edited by G Eysenbach; submitted 02.09.13; peer-reviewed by C Pfeiffer, C Broussard; comments to author 23.11.13; revised version received 09.01.14; accepted 12.01.14; published 20.02.14.

Please cite as:
Hämeen-Anttila K, Nordeng H, Kokki E, Jyrkkä J, Lupattelli A, Vainio K, Enlund H
Multiple Information Sources and Consequences of Conflicting Information About Medicine Use During Pregnancy: A Multinational Internet-Based Survey
J Med Internet Res 2014;16(2):e60
URL: http://www.jmir.org/2014/2/e60/
doi:10.2196/jmir.2939
PMID:24565696

Original Paper

Hospital Website Rankings in the United States: Expanding Benchmarks and Standards for Effective Consumer Engagement

Timothy R Huerta[1], MS, PhD; Jennifer L Hefner[2], MPH, PhD; Eric W Ford[3], MPH, PhD; Ann Scheck McAlearney[2], MS, ScD; Nir Menachemi[4], MPH, PhD

[1]Departments of Family Medicine and Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, United States
[2]Department of Family Medicine, College of Medicine, The Ohio State University, Columbus, OH, United States
[3]Bryan School of Business and Economics, University of North Carolina Greensboro, Greensboro, NC, United States
[4]Department of Health Care Organization and Policy, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Timothy R Huerta, MS, PhD
Departments of Family Medicine and Biomedical Informatics
College of Medicine, The Ohio State University
261 Northwood-High Building
2231 North High Street
Columbus, OH, 43201
United States
Phone: 1 614 366 6636
Fax: 1 614 293 7215
Email: Timothy.Huerta@osumc.edu

Abstract

Background: Passage of the Patient Protection and Affordable Care Act (ACA) increased the roles hospitals and health systems play in care delivery and led to a wave of consolidation of medical groups and hospitals. As such, the traditional patient interaction with an independent medical provider is becoming far less common, replaced by frequent interactions with integrated medical groups and health systems. It is thus increasingly important for these organizations to have an effective social media presence. Moreover, in the age of the informed consumer, patients desire a readily accessible, electronic interface to initiate contact, making a well-designed website and social media strategy critical features of the modern health care organization.

Objective: The purpose of this study was to assess the Web presence of hospitals and their health systems on five dimensions: accessibility, content, marketing, technology, and usability. In addition, an overall ranking was calculated to identify the top 100 hospital and health system websites.

Methods: A total of 2407 unique Web domains covering 2785 hospital facilities or their parent organizations were identified and matched against the 2009 American Hospital Association (AHA) Annual Survey. This is a four-fold improvement in prior research and represents what the authors believe to be a census assessment of the online presence of US hospitals and their health systems. Each of the five dimensions was investigated with an automated content analysis using a suite of tools. Scores on the dimensions are reported on a range from 0 to 10, with a higher score on any given dimension representing better comparative performance. Rankings on each dimension and an average ranking are provided for the top 100 hospitals.

Results: The mean score on the usability dimension, meant to rate overall website quality, was 5.16 (SD 1.43), with the highest score of 8 shared by only 5 hospitals. Mean scores on other dimensions were between 4.43 (SD 2.19) and 6.49 (SD 0.96). Based on these scores, rank order calculations for the top 100 websites are presented. Additionally, a link to raw data, including AHA ID, is provided to enable researchers and practitioners the ability to further explore relationships to other dynamics in health care.

Conclusions: This census assessment of US hospitals and their health systems provides a clear indication of the state of the sector. While stakeholder engagement is core to most discussions of the role that hospitals must play in relation to communities, management of an online presence has not been recognized as a core competency fundamental to care delivery. Yet, social media management and network engagement are skills that exist at the confluence of marketing and technical prowess. This paper presents performance guidelines evaluated against best-demonstrated practice or independent standards to facilitate improvement of the sector's use of websites and social media.

(J Med Internet Res 2014;16(2):e64) doi:10.2196/jmir.3054

KEYWORDS

social media; hospitals; information services; communication; access; consumer health information

Introduction

More than 80% of adults reported using Internet resources to support health care decisions in 2011 [1,2]. Thus, in many instances, a hospital's home page is the first point of contact for consumers [3,4]. As a result, the website for a hospital or health system has become an important communication channel for marketing to current and potential customers, as well as to visitors accompanying a patient [5-7]. Customers' evaluations of a hospital's website, and by extension their perceptions of the facility itself, will be based in part on comparisons to their experiences using other consumer websites such as Amazon and eBay [8,9]. If a hospital's website does not conform to or exceed a customer's expectations based on their previous experiences, negative inferences about facility quality will influence the decision-making process [10].

Reliance on search engines has resulted in individuals taking varied routes to their website destinations. Weaver et al [11] found that people seeking illness-related information behave differently than those seeking wellness information. Based on these trends, many hospital or health system websites have begun to include tools and information for patients and visitors that make navigating complex health encounters more user-friendly and that create a positive organizational image [12]. In so doing, hospitals are increasingly seeking to take on the role of trusted adviser, a role that is closely aligned with the accountable care organization (ACO) model in which health care providers work to empower patients to improve population health [7,13,14].

The competitive advantage gained from building an effective Web presence has led researchers to establish accessibility, content, marketing, and technical standards that define best demonstrated practices in website design [15-17]. In addition, the Health Information Technology Institute has issued standards for health care websites, including credibility, content, disclosure, links, design, interactivity, and caveats [18]. As a result, an exploration of US hospital and affiliated health system websites against the design standards used in other commercial endeavors is warranted to establish the state of the field.

The purpose of this article is to identify the degree to which hospitals and their health system websites comply with Internet-industry standards for commercial usability [19]. Using an automated Web crawler, we evaluated four dimensions—accessibility, content, marketing, and technology—using weighted multi-item scales. In addition, we used a weighted composite overall score to measure each website's quality across all four dimensions. The authors believe this analysis to be a census assessment for the online presence of hospitals and their health systems in the United States.

For hospital and health system decision makers, these analyses provide quantifiably objective, and immediately actionable, recommendations for enhancing the quality of their organization's website. Compared to other health information technology (HIT) upgrades that are being made to meet meaningful use goals set by the federal government, the capital investments required to create a state-of-the-art website are relatively modest and immediately visible. Additionally, as outlined above, having an effective website can create a competitive advantage when attracting consumers.

There is a dearth of evaluation studies on hospital websites. A number of factors play into this issue. Much of the research on the role of the Web in health include issues of accuracy, understandability, provenance, and timeliness. Williams et al [20] discuss a typology offered by Trochim on the purpose of websites and note that while the organizations don't keep their purposes neatly within the lines, they can still be assessed on the *raison d'être*—their reason for being and the services they aspire to provide. As such, there have been some efforts to study the content of health sites on a more specific basis—by judging the quality of the information as opposed to the technical merits of the system that is used for conveying that material. The American Public Health Association (APHA) offers a comprehensive statement on these issues [21].

First, while numerous studies exist focused on health information, hospitals are generally not seen as a portal for that purpose. This can be seen in the samples from which many of the health assessments are drawn—generally sites like WebMD, which seek to provide health information [20]—as opposed to local hospitals. While some facilities are large enough to have a national draw, and therefore serve the former purpose (eg, the Mayo Clinic, the Cleveland Clinic), most hospitals do not seek to provide health information, per se. Rather, they seek to provide information on services provided at the facility. As a result, hospitals serve a different role in the community, which generally is the delivery of health care services. This relationship is transforming as a result of the move to the ACO model in the United States, where hospitals are taking a greater responsibility for the care of the panels of patients they serve.

Ivory [22] notes that website evaluation is a moving target, suggesting that early assessments of evaluation focused on technical assessments of speed, while those have shifted over time. While some scholars have noted that human participation is fundamental to evaluating the usability of Web content, to ignore the value of automated assessment is to ignore the variability of experience provided to the end user by virtue of a site's failure to adhere to standards. Chiang and Starren [23] detail an assessment of consumer health website accessibility by users with sensory and physical disabilities, noting that W3C compliance is a minimum standard for ensuring accessibility, and adherence to that standard protocol should be framed as foundational and a precursor to more detailed assessment.

Methods

Overview

The names, cities, and states for every "non-government, not-for-profit (NFP)" or "investor-owned, for-profit" general medical or surgical hospital listed in the AHA 2009 Annual Survey in the United States (N=3523) were used in both Google and Bing to identify the first three returns for the search engines. These links were inspected to identify whether a hospital matching the facility of interest could be identified. In cases where a matching facility could not be found, an additional manual search was conducted in Google to attempt to locate the facility. Web site domains were included if and only if they could be attributed to the facility or a parent organization that was also a health services delivery organization. For instance, several critical access hospitals were excluded because they maintained a Web presence under a county government's Web domain. Website testing took place during the month of May 2013.

It should be noted that some organizations may have been a member of a health system, but also maintained their own domain. For instance, Northeast Arkansas Baptist Health System (neabaptist.com) is a hospital associated with Baptist Health System (baptistonline.org). In cases where a facility had its own domain, we assessed that domain separate and apart from the network or health system of which it was a part. The website of each organization was secured using a custom-built Web crawler. The Web crawler begins at the top-level Web page for the domain of each facility or system (eg, for the Kaiser Permanente domain, the Web crawler starts at the home page, kaiserpermanente.org), and drills down into successive subpages to build a topographical map of the links within a site. The analytic engine then samples 500 of these subpages and evaluates them based on a battery of assessment items, discussed at length in subsequent sections. A few websites were not assessed by the Web crawler due to technical problems, including timeout due to slow webpages or Web servers, server-side page redirections, or missing or unavailable host names. Only pages residing within the identified facility domain were tested.

To create summarized scores of website performance, the analytic engine scored content along five dimensions: (1) accessibility, (2) content, (3) marketing, (4) technology, and (5) usability. The scores on each dimension are reported on a range from 0 to 10, with a higher score on any given scale representing better comparative performance. The five dimensions detailed in the following sections provide broad assessments of aspects of website quality based on a set of underlying individual metrics. While it is important to note that some specific metrics contribute to more than one of the summarized scales, the scores themselves provide a basis for comparing two or more sites. The definitions of the specific items measured and how they are weighted in the summarized scores are presented in Multimedia Appendix 1.

Accessibility Dimension

Accessibility is a critical factor for reaching as many users as possible, but at-risk groups may not be familiar with access features that require higher levels of computer literacy, such as hovering over highlighted phrases to see additional information. Given the service domain in health care, the issue of accessibility is all the more important, and much discussion has been held on issues of access to services [24,25]. The accessibility score is an assessment of a website's ease of use for individuals with lower computer literacy levels, including those with physical disabilities that limit their use of a mouse or non-standard browser (such as mobile phones or tablet devices). A number of accessibility evaluation tools are offered by the World Wide Web Consortium (W3C) to explore these issues.

Content Dimension

The content dimension is an assessment of a website's overall content quality without taking into consideration the technical limitations of the site. Content quality is considered high if the text is grammatically correct, relevant, and updated regularly. The quality of the site's imagery (ie, photos and graphics) and metadata (ie, information about the data content in specific locations) is also assessed. Elements contributing to the content scale include individual tests of spelling, the degree to which the site adds new material, and the calculated reading age of the text on the pages. In particular, the Flesch-Kincaid readability metrics used in other health-related website studies are included as part of the content analysis [24,25]. The major measures that contribute to the content scale are freshness and the amount of content. The freshness measure is calculated by reading the dates that appear on a website's pages. Up-to-date content is a positive indicator to consumers that the organization is engaged in state-of-the-art activities. For example, monthly updates to the CEO's message may be understood to imply that a facility is customer-focused, while out-of-date content may foster a perception that public impressions are less important to the organization. Therefore, routinely adding and changing content to remain current and explicitly documenting the dates that Web pages are updated should be standard practice.

Marketing Dimension

The marketing dimension is an assessment of how readily and reliably information is accessed using search engines, including the appropriateness of content to hyperlinks, the rank and popularity of the website, and other technical aspects related to search engine optimization (SEO). SEO is an important aspect of the marketing scale. As content within a page becomes more accessible to search engines, the organization's profile in online searches becomes higher. Contributing individual tests include search engine results, search placement, and the use of content keywords that search engines rely on to prioritize websites. Performing these tasks effectively helps health systems maintain a consistent corporate image [26].

Technology Dimension

The technology dimension is an assessment of how well a website is designed, built, and maintained. Technical issues affect the user's experience and therefore can have a direct impact on the overall utility of the website for making decisions. Elements contributing to the technology index's scores include website download speed, site structure, code quality, and the use of cascading style sheets to organize content. The technology scale focuses purely on the performance aspects of a website without respect to its content. The major contributor to the score is the speed measure.

Usability Dimension

The usability score is a cross-sectional composite of a number of metrics used in other scales; therefore, it is a composite of metrics, not a composite of the other four scales. This dimension attempts to answer the question of how good a particular website is. Having this at-a-glance metric that rates the overall quality of a website as a single number enables comparisons across a number of critical areas of site presentation. The analytic engine also provides clear information about how each individual organization performs and, by extension, offers clues as to how improvements in these scores might be made.

Average Ranking Across Dimensions

Rank order calculations for each of the five dimensions were averaged to create a single average rank. The average rank score was then calculated across all website domains in order to rank hospitals from 1 to 2407.

Results

This search produced websites associated with 2407 unique domain names. Of those Web domains, 378 were attached to multiple AHA identifiers, indicating that they were part of larger organizations (ie, members in a system). In these cases, the system's domain was tested once rather than isolating a single facility from its parent organization. In total, 2785 facilities were scored. Organizational characteristics of these facilities are presented in Table 1.

Histograms representing the distribution of observations in each of the summarized scores are presented in Figures 1-5. Across all Figures, we find the scores to have a single mode with skewed distributions. The greatest variance can be seen in Figure 4, Technology. The mean scores for each dimension are presented in Table 2. Looking at the usability scale, the mean score was 5.16 (SD 1.43), with a maximum score of 8 achieved by only five organizations. Mean scores on the other scales ranged between 4.43 (SD 2.19) for technology and 6.49 (SD 0.96) for content.

The top 100 websites on each dimension are presented in Multimedia Appendix 2 with their respective rankings. The last column is an average rank score calculation across the five scales. For instance, jaxhealth.com scored 15th in Accessibility, 61st in Content, 52nd in Marketing, 14st in Technology, and 6th in Usability—resulting in an average score of 29.6, and making it the 1st best site overall. These rank order calculations, along with summary scores across dimensions, are presented for all 2407 websites in Multimedia Appendix 3. Included in these raw data is the AHA ID for each website domain.

Table 1. Hospital characteristics for all US AHA[a] hospitals by inclusion in the study.[b]

Hospital characteristics	Matched	Not-matched	Total US AHA[b] hospitals
Count of AHA IDs	2785.00	738.00	3523.00
Number of births	899.73	1035.24	928.12
Adjusted patient days	79,030.26	82,914.00	79,843.82
Transfer-adjusted admissions	15,299.73	16,595.29	15,571.13
Total expenditures	152,355,799.64	160,273,331.85	154,014,368.69
FTE employees	1011.65	1088.71	1027.80
Number of surgical operations	6387.38	7297.25	6577.98
Total visits	146,552.48	149,520.98	147,174.32
Number of beds	181.04	191.52	183.24
Average daily census	116.40	125.38	118.28

[a] AHA: American Hospital Association.

[b] Not-Matched Hospitals were excluded from the study because they did not have an identifiable Web domain attached to their name.

Table 2. Summary statistics for scales (n=2407).

Variable	Mean (SD)	Min	Max
Accessibility	5.08 (2.22)	0.0	9.0
Content	6.49 (0.96)	0.0	8.6
Marketing	5.03 (1.33)	0.8	8.5
Technology	4.43 (2.19)	0.0	8.7
Usability	5.16 (1.43)	0.0	8.0

Figure 1. Distribution of scores: Accessibility.

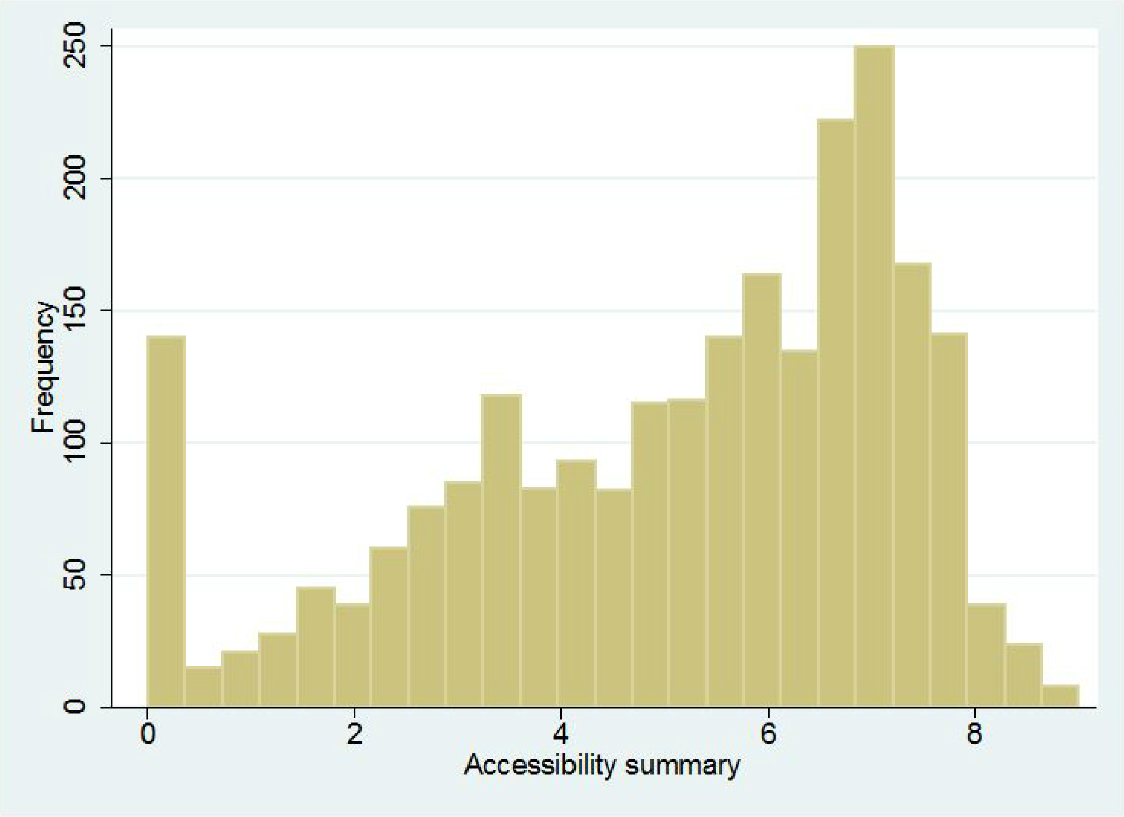


Figure 2. Distribution of scores: Content.

Figure 3. Distribution of scores: Marketing.

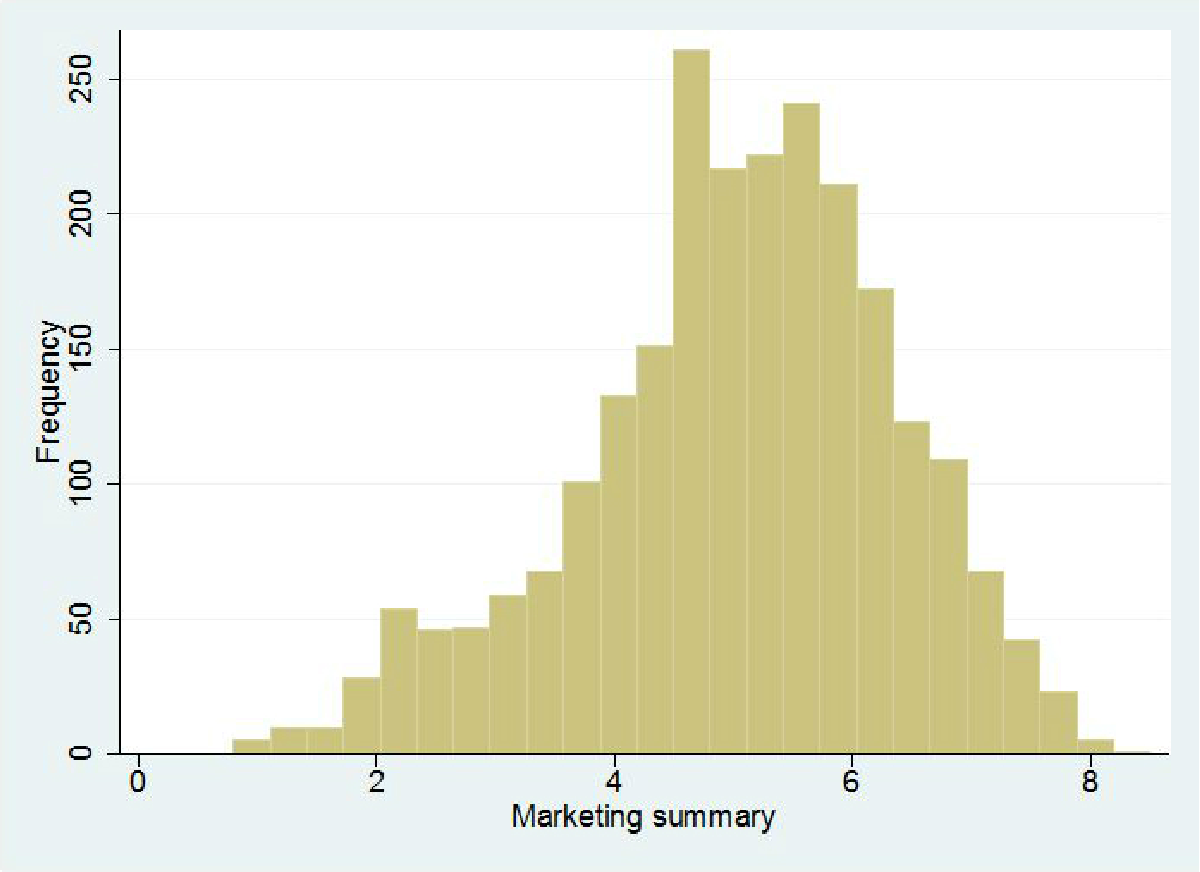


Figure 4. Distribution of scores: Technology.

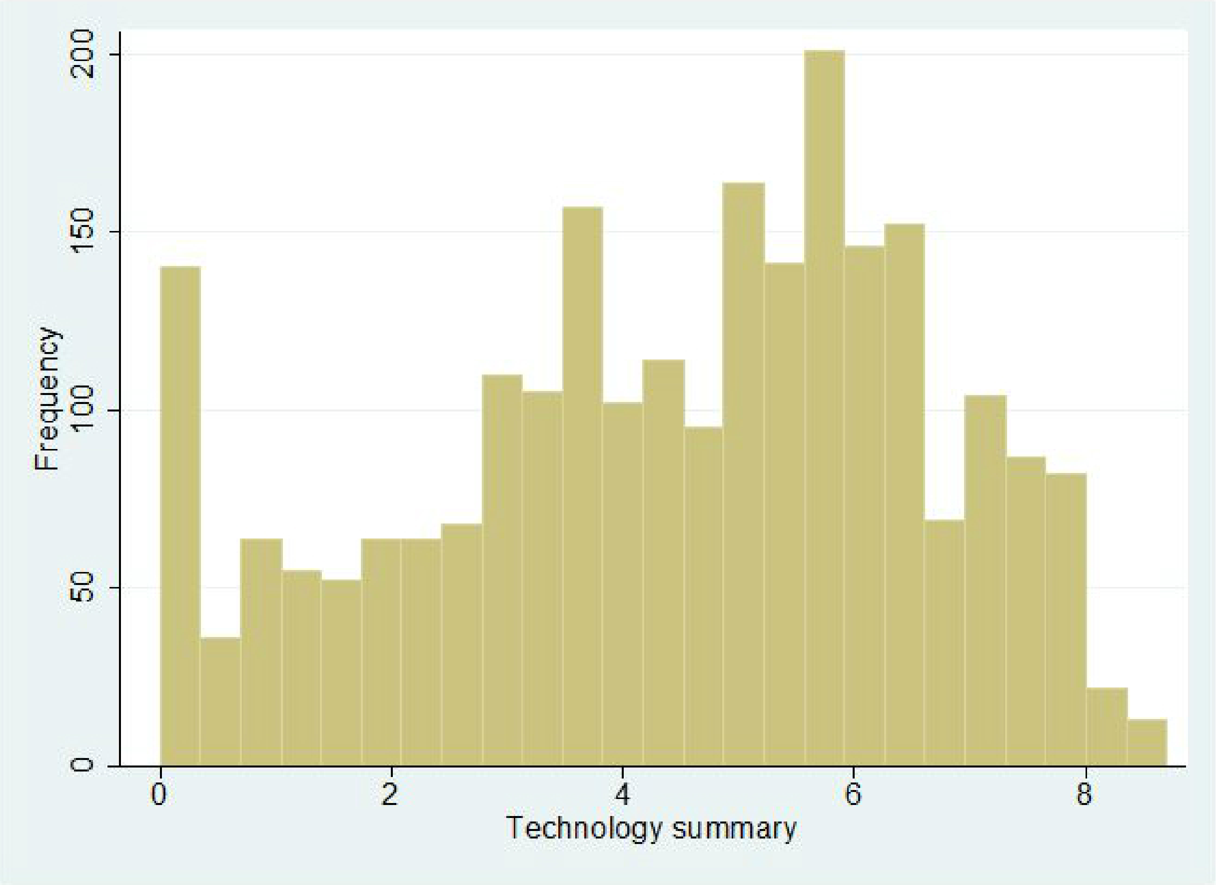
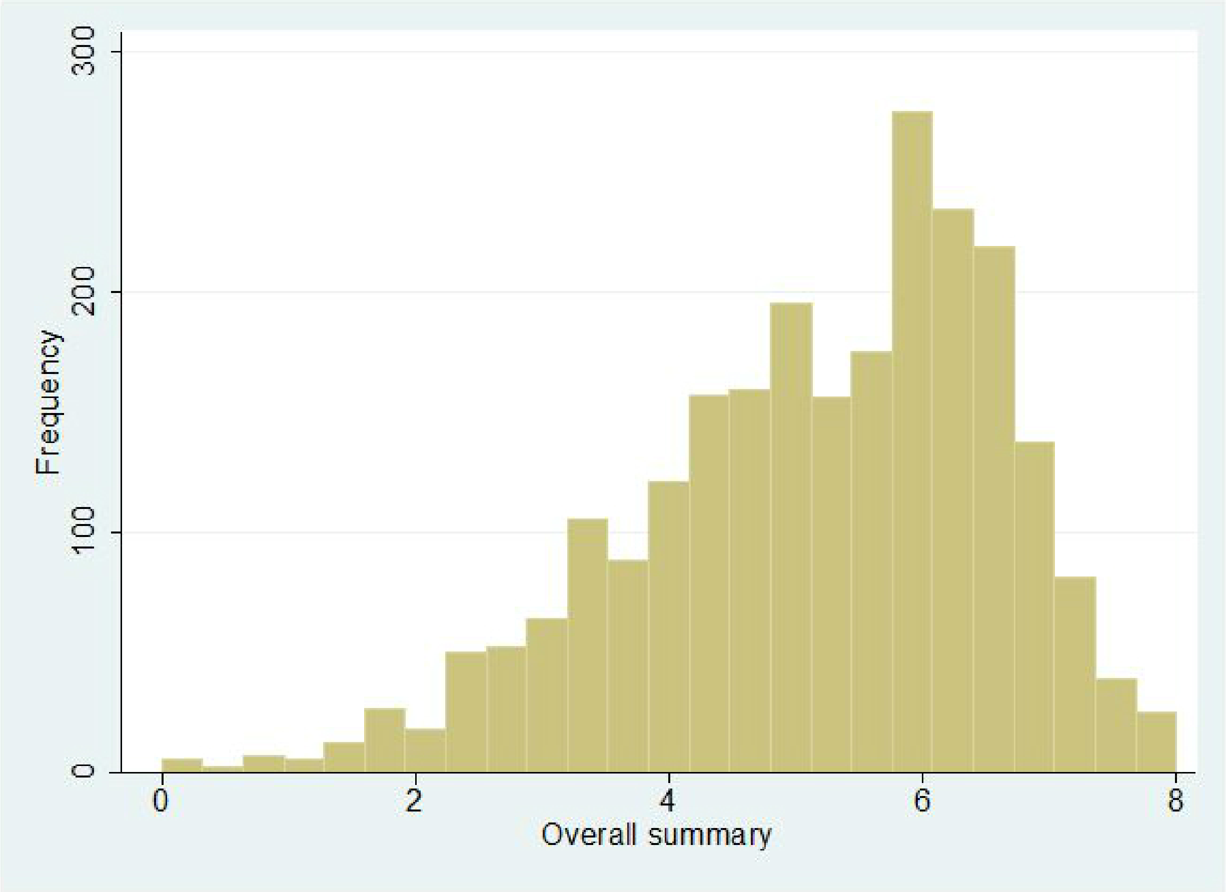


Figure 5. Distribution of scores: Overall.



Discussion

Principal Results and Implications

Those with an interest in evidence-based management know that one limitation is the presence of data upon which to make management decisions. While a website might have engaging photos and an interesting design, the proper functioning of a website requires far more work “under the hood” than most consider [27]. Layer upon that the importance of a broader social media strategy, and it can be easy to see how a facility might judge their online presence by the attractiveness of the pages. However, health care bears a special burden. With populations that are seeking care when they are least able to deal with externalities, a well-designed site that adheres to national standards and demonstrated best practices is important for issues related to care access [28,29]. Further, the absence of a comprehensive assessment across the entire sector could pose a significant detriment to care, given assumptions that may be made about accessibility and other factors. A thorough assessment requires the auditing of websites beyond the first pages. The present project, systematically assessing the website quality of 2785 hospitals, is a four-fold improvement over prior research and represents what the authors believe to be a census assessment for the online presence of US hospitals and their health systems.

The low mean score we found on the usability scale (5.16 on a 10-point scale, SD 1.43), a measure of general website quality, indicates that organizations’ websites, on average, have significant potential for improvement. In order to make a complete and effective assessment of a health care organization’s website, it is necessary to have the site evaluated using a Web crawling and analytic engine similar to the one employed in the present study. Nevertheless, there are many contributing scale components that organizational leaders can assess by a simple visual inspection of their website (eg, see the 2012 paper by Ford et al [30] for a more comprehensive discussion of this issue). Multimedia Appendix 1 presents a list of scale components with accompanying definitions and the percentage weight of each component within the five scales. These component definitions provide guidance for leaders to identify “low-hanging fruit” to improve their website scores. For example, one could easily search for and repair “Broken Links” (links to Web addresses that do not exist or return an error). Additionally, there is a “Spelling” component that assesses whether the words on a page are spelled correctly, and a “Twitter” component that determines if the website is linked to a Twitter account and how often tweets are posted.

While we did not test for the relationship between organizational type and website performance, a cursory review of the top 100 websites reveals that large facilities performed well. Consumers in urban areas typically go to a local facility for routine care or common procedures, but customers in rural areas may bypass their local facilities. In particular, consumers may travel for specialty care that is inaccessible locally, shopping for a facility and turning to the Internet for information [31,32]. Another explanation for large organizations’ success is that the complex and high-risk nature of cancer and childhood illnesses makes selecting an organization with high-quality ratings and a reputation for innovation particularly important [33,34]. This may contribute to the impetus for specialty facilities to develop high-performing websites for their marketing purposes.

For policymakers, the present analyses of health systems' websites may provide an indication of whether or not a health system is striving to become an ACO. Given the scope of organizational change required to become an ACO, it stands to reason that a health system's website would document and reflect such efforts in order to take advantage of that effort in the market. As a result, website quality may be a simple and reliable leading indicator of efforts to make this critical organizational transition. Prior research has found these dimensions of website quality to be linked to important aspects of patient care, including patient safety metrics [35], thus confirming the potential importance of our findings.

Limitations

The authors note several limitations to this study. Facilities associated with an education top domain (.edu) were purposely excluded from the analysis. The decision to exclude the .edu-based hospitals was a difficult one. While these hospitals represent a significant type of player in hospital health care delivery, they also often contain an academic side that would skew assessment. It was therefore impossible to create decision rules to exclude pages that were not directly tied to patient care or the hospital. As a result, inclusion of .edu-based content might result in academic departments unrelated to care influencing measures. It should also be noted that in some instances, hospitals create a facility domain that is separate and apart from the academic center. In these cases, where the domain is .org or .com, we assessed that content, but excluded .edu content if it was linked on any page. Therefore, a facility might have their clinic included, but the information about their doctors residing on the academic side might be excluded. The study authors intend to assess this subgroup in future analyses.

A similar dynamic occurred, at times, with select websites purchased outside of the .com and .org domains. For instance, facilities have also registered on the ".info" and ".us" domains. In these cases, the authors attempted to determine if the identified site was owned by the facility. These were judgment calls on the part of the research team, potentially resulting in counting errors, which would manifest in the descriptive statistics.

Another limitation of the research is related to the emerging nature of health care facilities. With the contraction and centralization of health care, there are fewer independent hospitals. Increasingly hospitals are joining networks and systems of care. These dynamics are becoming more pronounced in the wake of ACO development efforts. As a result, we can increasingly expect that access to regional information will be moved further down into Web pages as these systems centralize their marketing and information dissemination functions. Additionally, it should be noted that many facilities keep their content in a secure environment for users of their health systems. As a result, the assessment might incorrectly assess their Web presence. The assessment only reaches what is not secured based on links accessible from the home pages. This would simulate the information available to either non-members or members using only publically available information.

Finally, the authors recognize a concern around size and scope and the potential to misrepresent a single score as sufficient explanation of information given the diverse nature of facilities. In some cases, a website for a facility can be 6 pages. In another case, the system's website is over 10,000 pages and centrally managed. To then say that a domain is scored at 5.4 overall for both does not, in fact, mean they are equivalent. There are tradeoffs that any single measure must make. As a result, we have chosen to publish not only the overall score, but also the other tailored scores. The result is a greater nuance, but moves away from that single measure.

Comparison With Prior Work

In 2011, the websites of 636 US hospitals and health systems were tested using a similar methodology, with an overall mean score of 6.37 [30]. While we acknowledge that this prior work displayed higher scores, the addition of so many new facilities makes comparisons to prior evaluations problematic. The present study assessed 2407 domains that covered 2785 facilities. This is a four-fold improvement in prior research and represents what the authors believe to be a census assessment for the online presence of US hospitals and their health systems. We expect smaller facilities to have lower scores, on average. Put another way, we expect a lowering of scores, in comparison to scores calculated from prior research, as a result of greater inclusivity in our current assessment. Future iterations of this study will permit greater comparability across years.

Conclusions

The current analysis presents a significant update to the systematic assessment of hospital social media presence. Given the movement toward having health systems serve as ACOs that can empower consumers [14,30], the number of poorly performing facilities across all the calculated scores is concerning in the near term. The social media and Web presence of many of these organizations represents the first contact health care consumers make with the organization. If such contact fails to make a positive impression on the consumer, alternatives may be explored. In saturated markets where several organizations' services are interchangeable, a strong and well-designed Web and social media presence can be the difference between patients taking the first step into a facility or doing everything they can to avoid it. Health organizations should strive to standardize the quality of information presented on their websites [36], but they should also take care to deal with issues of accessibility, standards compliance, and search engine optimization.

Conflicts of Interest

The authors are faculty employed by their respective universities, some of which maintain a Web presence that was assessed in this study.

Multimedia Appendix 1

Scale components and weightings.

[PDF File (Adobe PDF File), 285KB - jmir_v16i2e64_app1.pdf]

Multimedia Appendix 2

Ranking of the Top 100 websites for each dimension and an average ranking across dimensions.

[PDF File (Adobe PDF File), 394KB - jmir_v16i2e64_app2.pdf]

Multimedia Appendix 3

Complete dataset including domains, associated composite scores, and rankings. (Crosswalk data provided in separate sheet needs to be merged to use the dataset. Not provided merged due to a concern that univariate analysis of the merged data will multiply count single observations, because facilities may share a domain).

[XLSX File (Microsoft Excel File), 469KB - jmir_v16i2e64_app3.xlsx]

References

1. Reid P, Borycki EM. Emergence of a new consumer health informatics framework: introducing the healthcare organization. Stud Health Technol Inform 2011;164:353-357. [Medline: 21335736]
2. Szokan N. The Washington Post. 2011 Feb 01. Health information remains high on the list of popular uses for the Internet URL: http://www.washingtonpost.com/wp-dyn/content/article/2011/01/31/AR2011013104786.html [accessed 2014-02-16] [WebCite Cache ID 6NRM3vimj]
3. Alpay L, Overberg RI, Schonk BZ. Empowering citizens in assessing health related websites: A driving factor for healthcare governance. International Journal of Healthcare Technology & Management 2007 Jan 01;8(1):141-160 [FREE Full text]
4. Revere L, Robinson L. How healthcare organizations use the Internet to market quality achievements. J Healthc Manag 2010;55(1):39-49; discussion 49. [Medline: 20210072]
5. Coile RC. E-health: reinventing healthcare in the information age. J Healthc Manag 2000;45(3):206-210. [Medline: 11066968]
6. Randeree E, Rao HR. E-health and assurance: curing hospital websites. Int J Electron Healthc 2004;1(1):33-46. [Medline: 18048202]
7. Wen KY, Kreps G, Zhu F, Miller S. Consumers' perceptions about and use of the Internet for personal health records and health information exchange: analysis of the 2007 Health Information National Trends Survey. J Med Internet Res 2010;12(4):e73 [FREE Full text] [doi: 10.2196/jmir.1668] [Medline: 21169163]
8. Mittler J, Volmar KM, Shaw BW, Christianson JB, Scanlon DP. Using websites to engage consumers in managing their health and healthcare. Am J Manag Care 2012 Sep;18(6 Suppl):s177-s184 [FREE Full text] [Medline: 23286713]
9. Liang C, Chen H. A study of the impacts of website quality on customer relationship performance. Total Quality Management & Business Excellence 2009;20(9):971-988.
10. Rutitis D, Batraga A, Muizniece L, Ritovs K. Management of corporate identity dimensions in the health care. Procedia - Social and Behavioral Sciences 2012 Oct;58:995-1003. [doi: 10.1016/j.sbspro.2012.09.1079]
11. Weaver JB, Mays D, Weaver SS, Hopkins GL, Eroglu D, Bernhardt JM. Health information-seeking behaviors, health indicators, and health risks. Am J Public Health 2010 Aug;100(8):1520-1525. [doi: 10.2105/AJPH.2009.180521] [Medline: 20558794]
12. LaPenna AM. New sources of revenue in the age of connectivity. J Healthc Manag 2009;54(1):11-14. [Medline: 19227850]
13. O'Donnell J, Hannan EL. An analysis of case-intensity-adjusted nursing staff patterns in New York State skilled nursing facilities. Med Care 1983 Jan;21(1):82-91. [Medline: 6403784]
14. Simborg DW. Consumer empowerment versus consumer populism in healthcare IT. J Am Med Inform Assoc 2010;17(4):370-372 [FREE Full text] [doi: 10.1136/jamia.2010.003392] [Medline: 20595301]
15. Oermann MH, McInerney SM. An evaluation of sepsis Web sites for patient and family education. Plast Surg Nurs 2007;27(4):192-196. [doi: 10.1097/01.PSN.0000306184.95812.8a] [Medline: 18165726]
16. Oermann MH, Lesley ML, VanderWal JS. Using Web sites on quality health care for teaching consumers in public libraries. Qual Manag Health Care 2005;14(3):188-195. [Medline: 16027597]
17. Oermann MH, Lowery NF, Thornley J. Evaluation of Web sites on management of pain in children. Pain Manag Nurs 2003 Sep;4(3):99-105. [Medline: 14566707]

18. The Agency for Healthcare Research and Quality. Assessing the Quality of Internet Health Information: Summary. 1999. URL: http://www.ahrq.gov/research/data/infoqual.html [accessed 2014-02-16] [WebCite Cache ID 6NRMyDsNL]
19. W3C Standards. 2013. URL: http://www.w3.org/standards/ [accessed 2014-02-16] [WebCite Cache ID 6NRN24rFN]
20. Williams P, Nicholas D, Huntington P, McLean F. Surfing for health: user evaluation of a health information website. Part one: Background and literature review. Health Info Libr J 2002 Jun;19(2):98-108. [Medline: 12389607]
21. Criteria for Assessing the Quality of Health Information on the Internet. 2002. URL: http://www.apha.org/NR/exeres/412DDD11-42A0-46CB-8705-F9D47A073AF1.htm [accessed 2014-02-16] [WebCite Cache ID 6NRN56Hqx]
22. Ivory MY. Automated Web site evaluation: researchers' and practitioners' perspectives. In: Human-Computer Interaction Series. The Netherlands: Kluwer Academic Publishers; 2003.
23. Chiang MF, Starren J. Evaluation of consumer health website accessibility by users with sensory and physical disabilities. Stud Health Technol Inform 2004;107(Pt 2):1128-1132. [Medline: 15360988]
24. Stewart MJ, Georgiou A, Westbrook JI. Successfully integrating aged care services: a review of the evidence and tools emerging from a long-term care program. Int J Integr Care 2013;13:e003 [FREE Full text] [Medline: 23687475]
25. Clarke S, Calam R, Morawska A, Sanders M. Developing web-based Triple P 'Positive Parenting Programme' for families of children with asthma. Child Care Health Dev 2013 May 12:1. [doi: 10.1111/cch.12073] [Medline: 23662595]
26. Batraga A, Rutitis D. Corporate identity with the health care industry. Economics and Management 2012 Dec 07;17(4):1545-1551. [doi: 10.5755/j01.em.17.4.3027]
27. Griffiths KM, Christensen H. Website quality indicators for consumers. J Med Internet Res 2005;7(5):e55 [FREE Full text] [doi: 10.2196/jmir.7.5.e55] [Medline: 16403719]
28. 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]
29. Zeng X, Parmanto B. Web content accessibility of consumer health information Web sites for people with disabilities: a cross sectional evaluation. J Med Internet Res 2004 Jun 21;6(2):e19 [FREE Full text] [doi: 10.2196/jmir.6.2.e19] [Medline: 15249268]
30. Ford EW, Huerta TR, Schilhavy RA, Menachemi N. Effective US health system websites: establishing benchmarks and standards for effective consumer engagement. J Healthc Manag 2012;57(1):47-64; discussion 64. [Medline: 22397104]
31. Saunders C, Bellamy GR, Menachemi N, Chukmaitov AS, Brooks RG. Bypassing the local rural hospital for outpatient procedures. J Rural Health 2009;25(2):174-181. [doi: 10.1111/j.1748-0361.2009.00214.x] [Medline: 19785583]
32. Damman OC, van den Hengel YK, van Loon AJ, Rademakers J. An international comparison of web-based reporting about health care quality: content analysis. J Med Internet Res 2010;12(2):e8 [FREE Full text] [doi: 10.2196/jmir.1191] [Medline: 20439252]
33. Kim K, Kwon N. Profile of e-patients: analysis of their cancer information-seeking from a national survey. J Health Commun 2010 Oct;15(7):712-733. [doi: 10.1080/10810730.2010.514031] [Medline: 21104502]
34. Kinnane N. Evaluation of a hospital-based cancer information and support centre. Support Care Cancer 2012 Feb;20(2):287-300. [doi: 10.1007/s00520-010-1071-y] [Medline: 21225290]
35. Ford E, Huerta TR, Schilhavy RA, Menachemi N. Effective US health system websites: establishing benchmarks and standards for effective consumer engagement. J Healthc Manag 2012;57(1):47-64; discussion 64. [Medline: 22397104]
36. Suchy K. A lack of standardization: the basis for the ethical issues surrounding quality and performance reports. J Healthc Manag 2010;55(4):241-251. [Medline: 20812525]

Abbreviations

ACO: accountable care organization
AHA: American Hospital Association
SEO: search engine optimization

Edited by G Eysenbach; submitted 28.10.13; peer-reviewed by A Kazley, A Vagelatos; comments to author 23.12.13; revised version received 26.12.13; accepted 09.01.14; published 25.02.14.

Please cite as:
Huerta TR, Hefner JL, Ford EW, McAlearney AS, Menachemi N
Hospital Website Rankings in the United States: Expanding Benchmarks and Standards for Effective Consumer Engagement
J Med Internet Res 2014;16(2):e64
URL: http://www.jmir.org/2014/2/e64/
doi:10.2196/jmir.3054
PMID:24568892

Original Paper

Validation and Reliability of a Smartphone Application for the International Prostate Symptom Score Questionnaire: A Randomized Repeated Measures Crossover Study

Jae Heon Kim[1*], MD, PhD; Soon-Sun Kwon[2*], PhD; Sung Ryul Shim[3]; Hwa Yeon Sun[1]; Young Myoung Ko[4], PhD; Dong-Il Chun[5], MD, PhD; Won Jae Yang[1], MD, PhD; Yun Seob Song[1], MD, PhD

[1]Department of Urology, Soonchunhyang University, College of Medicine, Soonchunhyang University Hospital, Seoul, Korea, Republic Of
[2]Biomedical Research Institute, Seoul National University Bundang Hospital, Seongnam, Korea, Republic Of
[3]Department of Epidemiology and Medical Informatics, Korea University, Seoul, Korea, Republic Of
[4]Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Korea, Republic Of
[5]Department of Orthopaedics, Soonchunhyang University, College of Medicine, Soonchunhyang University Hospital, Seoul, Korea, Republic Of
[*]these authors contributed equally

Corresponding Author:
Yun Seob Song, MD, PhD
Department of Urology, Soonchunhyang University
College of Medicine
Soonchunhyang University Hospital
59, Daesagwan-ro, Yongsan-gu
Seoul, 140-743
Korea, Republic Of
Phone: 82 2 709 9377
Fax: 82 2 709 9378
Email: yssong@schmc.ac.kr

Abstract

Background: Smartphone-based assessment may be a useful diagnostic and monitoring tool for patients. There have been many attempts to create a smartphone diagnostic tool for clinical use in various medical fields but few have demonstrated scientific validity.

Objective: The purpose of this study was to develop a smartphone application of the International Prostate Symptom Score (IPSS) and to demonstrate its validity and reliability.

Methods: From June 2012 to May 2013, a total of 1581 male participants (≥40 years old), with or without lower urinary tract symptoms (LUTS), visited our urology clinic via the health improvement center at Soonchunhyang University Hospital (Republic of Korea) and were enrolled in this study. A randomized repeated measures crossover design was employed using a smartphone application of the IPSS and the conventional paper form of the IPSS. Paired *t* test under a hypothesis of non-inferior trial was conducted. For the reliability test, the intraclass correlation coefficient (ICC) was measured.

Results: The total score of the IPSS (P=.289) and each item of the IPSS (P=.157-1.000) showed no differences between the paper version and the smartphone version of the IPSS. The mild, moderate, and severe LUTS groups showed no differences between the two versions of the IPSS. A significant correlation was noted in the total group (ICC=.935, P<.001). The mild, moderate, and severe LUTS groups also showed significant correlations (ICC=.616, .549, and .548 respectively, all P<.001).There was selection bias in this study, as only participants who had smartphones could participate.

Conclusions: The validity and reliability of the smartphone application version were comparable to the conventional paper version of the IPSS. The smartphone application of the IPSS could be an effective method for measuring lower urinary tract symptoms.

(J Med Internet Res 2014;16(2):e38) doi:10.2196/jmir.3042

KEYWORDS

smartphone; International Prostate Symptom Score; lower urinary tract symptoms; health survey; questionnaires

Introduction

Lower urinary tract symptoms (LUTS) suggestive of benign prostatic hyperplasia (BPH) affect a majority of men and show an increasing prevalence with aging [1]. Male patients with LUTS suffer from significant impairment of their quality of life (QOL) and interference in daily living activities due to urinary dysfunction [2].

The primary treatment goal for men with clinical manifestations of BPH is to reduce or relieve LUTS. Therefore, the measurement of LUTS is a key factor in the evaluation of these patients, both in clinical practice and in research studies [2,3].

The International Prostate Symptom Score (IPSS) is the most widely used diagnostic tool in urology and is widely available, validated, and has been translated into many languages [3]. The IPSS scale correlates positively and significantly with global measures of difficulty and health complications associated with LUTS [4-6].

To date, the paper-based IPSS has been used worldwide. Although paper-based questionnaires have been the standard tools for screening or monitoring of medical conditions, this method has several problems, including data collection and entry errors [7].

Over the past 20 years, questionnaires have been developed using electronic systems, such as Web surveys on personal computers or personal digital assistants, and more recently using smartphones. There has been a meteoric rise in the use of smartphones, which has reached approximately 6 billion people worldwide [8], enabling smartphones to function as new tools for measuring the health of individuals.

Most smartphone applications have focused on education and communication for medical school students and clinicians [9]. However, several smartphone application questionnaires have been introduced in clinical use, including psychiatry and sleep disorders [10-12].Considering the worldwide use of the IPSS in clinical fields, a smartphone application of the IPSS could be very popular, both for patients and clinicians.

The ultimate goal of smartphone-based assessment is the establishment of a home diagnostic device that enables easy screening and monitoring of a disease by scoring data and thus reducing the time to diagnosis and treatment, as well as overall costs. Most questionnaires are originally designed as paper-based questionnaires and, therefore, validation of a smartphone-based version is required due to the possibility of response bias between paper and smartphone versions [13].The aim of this study was to examine the validity and reliability of a smartphone application version of the IPSS by quantitative analysis and to show the satisfaction rate compared with the conventional paper version of the IPSS.

Methods

Study Sample

From June 2012 to May 2013, 1581 male patients presenting with LUTS (≥ 40 years old) visited our urology clinic via the health improvement center at Soonchunhyang University Hospital (Republic of Korea) and were enrolled in this study. All patients underwent a complete history, physical examination, and urinalysis, and completed an IPSS questionnaire. Patients who had a history of cancer of any organ, neurologic diseases or disorders, uncontrolled hypertension or diabetes, psychiatric disorders, prostatic surgery, liver cirrhosis, or renal failure were excluded from this study. Participant data were recorded in a prospective database and the study was approved by the Institutional Review Board of Soonchunhyang University Hospital.

Methodology

A randomized repeated measures crossover design was employed using the smartphone application and conventional paper form of the IPSS. One investigator conducted face-to-face interviews with all study participants, using structured explanations. The Korean version of the IPSS has been verified for relevance and reliability, and is the most popular diagnostic instrument for LUTS in Korea [14]. There was a 1-week break before and after completion of the alternative versions of the IPSS in order to reduce carryover effects. Questionnaires were randomly assigned to the smartphone or paper version of the IPSS. The supervisors obtained oral informed consent from participants before the study.

Developing the Smartphone Application of the IPSS

The smartphone application was specifically developed for Android model smartphones (Android is the operating system created by Google). While actively answering the IPSS in the smartphone application, participants could go backward to correct answers before they chose “save” to go on to the next question or to finish the test. The responses on the smartphone application of the IPSS were automatically transferred to the database where only supervisors could access the information.

System Stability

For a successful study trial, the stability of the system providing the application of the IPSS should be safe and stable. To this end, the server manager closely monitored the system during the performance of this trial. Prior to this trial, we conducted a pilot test to check the stability of the smartphone application of the IPSS and the data collection server.

Main Outcome for Validity

The overall hypothesis of this study was that participants would find both the smartphone application and paper version of the IPSS feasible and acceptable to use, which means that the smartphone application of the IPSS would not be inferior to the paper version of the IPSS. Validity was defined by non-difference in the total score of the IPSS and in each item of the IPSS.

Main Outcome for Reliability

Reliability referred to the consistency of IPSS scores obtained by the same person between the paper and smartphone version. There were many statistics available to measure reliability. The reliability test was conducted by the intraclass correlation coefficient (ICC). ICC was used to measure reliability for the paper and smartphone IPSS scores and ranged from 0 (no agreement) to 1 (perfect agreement). Reliability was defined by correlation between the two versions of the IPSS.

Delivery Method Compliance

In order to determine the compliance rate, the participation and preference rates were investigated by specific questions at the end of the trial: “Which method would you be willing to use to complete the assessment more easily?” and “Which method would you prefer?”

Power Calculation

The base of our sample size was dependent on the hypothesis that a smartphone application of the IPSS was not inferior to the conventional paper-based IPSS [15]. We used an alpha error of .05 and beta error of .2. The calculated sample size was 980. Considering a 10% decline or withdrawal rate, the minimal sample size was calculated to be 1100.

Data Analysis

We analyzed differences by paired *t* test for the two pairs of questionnaires. Reliability was assessed using ICC and a two-way random effect model, assuming a single measurement and absolute agreement. Statistical analyses were performed using SPSS version 20.0 for Windows. All statistics were two-tailed and *P* values <.05 were considered statistically significant.

Preservation of Data Security

Individual or total scores were not accessible by participants through their smartphones. After completing the paper version of the IPSS, the data collector recorded the data after checking the data transfer error and then sent the data to the statistician. After completing the smartphone application of the IPSS, the data were transferred automatically to the special server and then sent to the statistician. After completing the analysis, all information regarding the IPSS was removed by the supervisor.

Results

Basic Characteristics of the Participants

The mean age of the participants was 58.49 (SD 7.22). The mean total score of the IPSS paper version was 11.04 (SD 7.76) and the mean total IPSS score of the smartphone version was 11.03 (SD 7.77). There were 668 (42.25%, 668/1581) subjects in the mild LUTS group, 643 (40.67%, 643/1581) subjects in the moderate LUTS group, and 270 (17.08%, 270/1581) subjects in the severe LUTS group (Table 1). The refusal rate in total was 16 and there were 3 cases of missing data, so the final allocation was 1581 cases (Figure 1).

The total rates of mild, moderate, and severe LUTS groups were significantly different (*P*<.001). In the comparison of the rates of these groups, the mild and moderate LUTS groups were not different (*P*=.471), but there were significant differences (*P*<.001) between the mild and severe LUTS groups and the moderate and severe LUTS groups.

Figure 1. Flow chart of study participants.

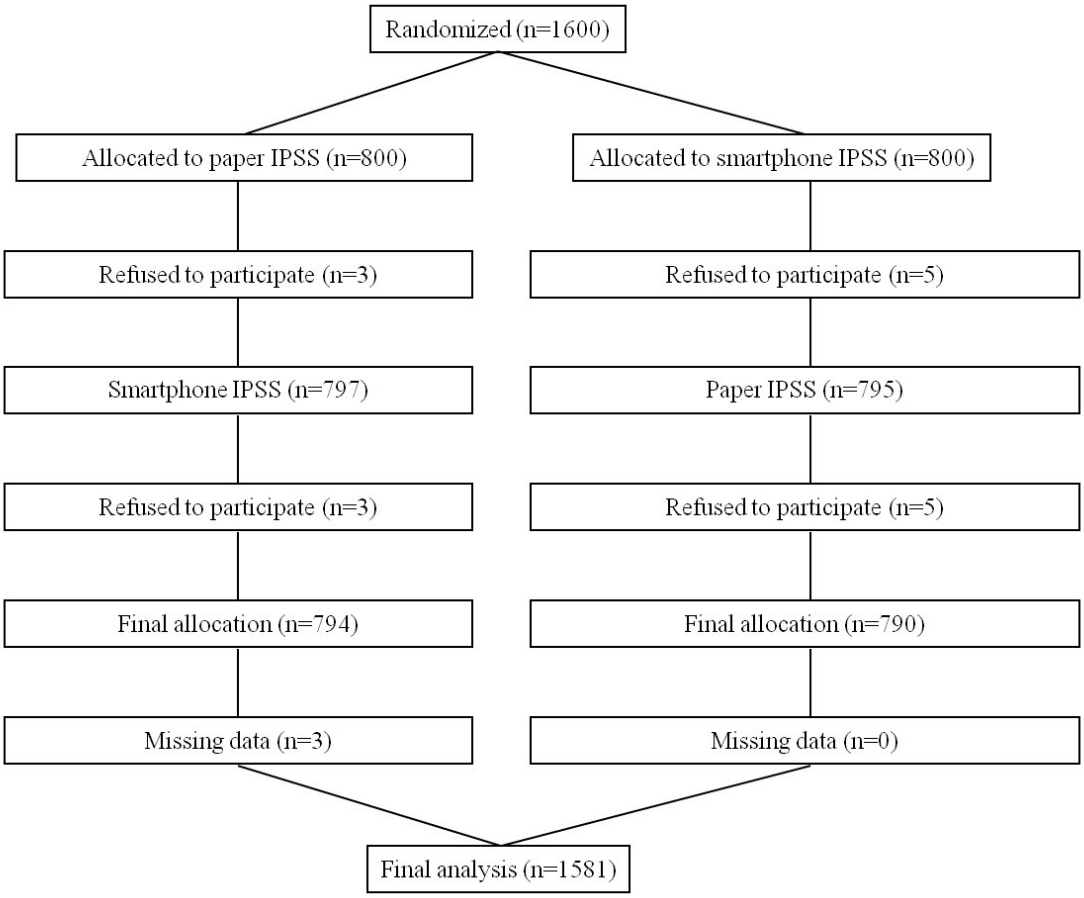


Table 1. Baseline characteristics (n=1581).

	n	Mean (SD)	Range
Age	1581	58.49 (7.22)	40-79
Paper IPSS[a] Q1	1580	1.69 (1.61)	0-5
Paper IPSS Q2	1581	1.57 (1.42)	0-5
Paper IPSS Q3	1579	1.63 (1.54)	0-5
Paper IPSS Q4	1579	1.32 (1.42)	0-5
Paper IPSS Q5	1581	2.30 (1.66)	0-5
Paper IPSS Q6	1580	1.27 (1.46)	0-5
Paper IPSS Q7	1581	1.28 (1.15)	0-7
Paper Total IPSS	1581	11.04 (7.77)	0-35
Paper QOL[b]	1546	2.77 (1.38)	0-6
Smartphone IPSS Q1	1580	1.68 (1.61)	0-5
Smartphone IPSS Q2	1581	1.57 (1.42)	0-5
Smartphone IPSS Q3	1579	1.63 (1.54)	0–5
Smartphone IPSS Q4	1579	1.32 (1.42)	0-5
Smartphone IPSS Q5	1581	2.30 (1.67)	0-5
Smartphone IPSS Q6	1580	1.27 (1.46)	0-5
Smartphone IPSS Q7	1581	1.28 (1.16)	0-7
Smartphone Total IPSS	1581	11.03 (7.77)	0-35
Smartphone QOL	1546	2.77 (1.37)	0-6
Mild LUTS[c]	668 (42.25%)	--	--
Moderate LUTS	643 (40.67%)	--	--
Severe LUTS	270 (17.08%)	--	--

[a] IPSS: International Prostate Symptom Score

[b] QOL: quality of life

[c] LUTS: lower urinary tract symptoms

Validation With Paired *t* Test

There were no differences in the overall total scores of the IPSS (*P*=.289) and each item of the IPSS (*P* values =.157-1.000) between the paper version of the IPSS and smartphone version of the IPSS (Table 2). In the mild LUTS group, the total score of the IPSS (*P*=.752) and each item of the IPSS (*P* values =.157-1.000) showed no differences between the paper version and the smartphone version of the IPSS (Table 3). In the moderate LUTS group, the total score of the IPSS (*P*=.432) and each item of the IPSS (*P* values =.103-1.000) showed no differences between the paper version and the smartphone version of the IPSS (Table 4). In the severe LUTS group, the total score of the IPSS (*P*=.083) and each item of the IPSS (*P* values =.158-1.000) showed no differences between the paper version and the smartphone version of the IPSS (Table 5).

Table 2. Paired *t* test between paper and smartphone IPSS[a].

	Paper, mean (SD)	Smartphone, mean (SD)	*P* value
IPSS Q1	1.69 (1.61)	1.68 (1.61)	.157
IPSS Q2	1.57 (1.42)	1.57 (1.42)	.166
IPSS Q3	1.63 (1.54)	1.63 (1.54)	.435
IPSS Q4	1.32 (1.42)	1.32 (1.42)	.578
IPSS Q5	2.30 (1.66)	2.30 (1.66)	.223
IPSS Q6	1.27 (1.46)	1.27 (1.46)	1.000
IPSS Q7	1.28 (1.15)	1.28 (1.16)	.317
Total IPSS	11.04 (7.77)	11.03 (7.77)	.289
IPSS QOL[b]	2.77 (1.38)	2.77 (1.37)	.180

[a] IPSS: International Prostate Symptom Score

[b] QOL: quality of life

Table 3. Paired *t* test between paper and smartphone IPSS[a] in mild LUTS[b].

	Paper, mean (SD)	Smartphone, mean (SD)	*P* value
IPSS Q1	0.53 (0.71)	0.53 (0.71)	1.000
IPSS Q2	0.60 (0.71)	0.60 (0.71)	.706
IPSS Q3	0.53 (0.67)	0.53 (0.66)	.157
IPSS Q4	0.39 (0.62)	0.40 (0.62)	.257
IPSS Q5	0.99 (0.92)	0.99 (0.93)	1.000
IPSS Q6	0.37 (0.62)	0.38 (0.620	.706
IPSS Q7	0.73 (0.77)	0.72 (0.77)	.180
Total IPSS	4.13 (2.09)	4.13 (2.10)	.752
IPSS QOL[c]	1.86 (1.10)	1.86 (1.10)	1.000

[a] IPSS: International Prostate Symptom Score

[b] LUTS: lower urinary tract symptoms

[c] QOL: quality of life

Table 4. Paired *t* test between paper and smartphone IPSS[a] in moderate LUTS[b].

	Paper, mean (SD)	Smartphone, mean (SD)	*P* value
IPSS Q1	1.93 (1.30)	1.93 (1.30)	.318
IPSS Q2	1.82 (1.14)	1.81 (1.14)	.103
IPSS Q3	1.87 (1.19)	1.87 (1.20)	.631
IPSS Q4	1.53 (1.19)	1.53 (1.19)	.655
IPSS Q5	2.75 (1.31)	2.74 (1.30)	.127
IPSS Q6	1.36 (1.15)	1.35 (1.15)	.819
IPSS Q7	1.34 (1.01)	1.34 (1.01)	1.000
Total IPSS	12.59 (3.30)	12.58 (3.32)	.432
IPSS QOL[c]	2.99 (1.02)	2.99 (1.01)	.318

[a] IPSS: International Prostate Symptom Score
[b] LUTS: lower urinary tract symptoms
[c] QOL: quality of life

Table 5. Paired *t* test between paper and smartphone IPSS[a] in severe LUTS[b].

	Paper, mean (SD)	Smartphone, mean (SD)	*P* value
IPSS Q1	3.96 (1.12)	3.95 (1.15)	.318
IPSS Q2	3.38 (1.33)	3.38 (1.33)	1.000
IPSS Q3	3.81 (1.26)	3.81 (1.26)	1.000
IPSS Q4	3.09 (1.49)	3.09 (1.49)	1.000
IPSS Q5	4.46 (0.79)	4.46 (0.79)	.158
IPSS Q6	3.29 (1.51)	3.29 (1.52)	1.000
IPSS Q7	2.51 (1.28)	2.51 (1.28)	1.000
Total IPSS	24.45 (3.58)	24.44 (3.58)	.083
IPSS QOL[c]	4.45 (0.81)	4.43 (0.82)	.318

[a] IPSS: International Prostate Symptom Score
[b] LUTS: lower urinary tract symptoms
[c] QOL: quality of life

Reliability Test With Interclass Correlation Coefficient (ICC) Test

Reliability was assessed using the ICC and a two-way random effect model, assuming a single measurement and absolute agreement. A significant correlation was noted in the total group (ICC=.935, *P*<.001). The mild, moderate, and severe LUTS groups also showed significant correlations (ICCs =.616, .549, and .548 respectively, all *P* values <.001) (Table 6).

Table 6. Reliability test of paper and smartphone versions of the IPSS[a].

	Interclass correlation coefficient	95% CI		*P* value
		Lower	Upper	
Total	.935	0.927	0.941	<.001
Mild LUTS[b]	.616	0.571	0.659	<.001
Moderate LUTS	.549	0.492	0.602	<.001
Severe LUTS	.548	0.462	0.625	<.001

[a] IPSS: International Prostate Symptom Score
[b] LUTS: lower urinary tract symptoms

Compliance

For compliance, we created two questions that asked, “Which method would you be willing to use to complete the assessment more easily?” and “Which method would you prefer?” In the examination of feasibility, 760 (48.07%, 760/1581) participants replied that the smartphone version was more feasible, 420 (26.56%, 420/1581) participants replied that the paper version was more feasible, and 301 (19.03%, 301/1581) participants replied that both of the versions were feasible. With regard to preference, 820 (51.86%, 820/1581) participants preferred the smartphone version, 320 participants (20.24%, 320/1581) preferred the paper version, and 356 (22.51%, 356/1581) participants showed no preference. For the two questionnaires, the results showed significant differences according to age (*P*<.001).

Discussion

Principal Findings

The objective of this study was to compare two different methods for diagnostic assessment via screening questionnaires. The overall hypothesis of this study was that a smartphone application of IPSS would show the same efficacy in real clinical use as a diagnostic questionnaire. This study was the first to show the clinical use of a smartphone questionnaire application in the field of urology (Figures 2-4).

The recent remarkable increase in the adoption of smartphones enables not only easy communication, but also the possibility of utilizing devices in diverse settings, including health care. There have been several studies using smartphone applications of sleep questionnaires, including the Epworth Sleepiness Scale, the Berlin questionnaire, and the STOP BANG questionnaire [10]. In psychiatric disease and serious mental illness, smartphone-based questionnaires for monitoring have shown better compliance than other types of questionnaires. Ambulatory monitoring of symptoms using smartphone applications represents a feasible and valid way of assessing psychotic status for research and clinical management [11,16]. Smartphone application questionnaires have also been used in a cardiac rehabilitation population for measuring physical activity with demonstrated validity and reliability [12].For clinical application as a medical device, a smartphone-based application has been used for actigraphy recording and audio recording in sleep disorders [10]. However, none were statistically validated by a comparison with the paper version of the questionnaire by a non-inferior trial with the hypothesis that the smartphone version of the questionnaire could be as effective as the paper version of the questionnaire. Little is known about the strengths and limitations of smartphone application questionnaires in the screening or monitoring of diseases, especially in the field of urology. Only one study has shown the validity and reliability of a smartphone application for the assessment of penile deformity in Peyronie’s disease [17]. There are several urologic applications such as “Bladder Pal”, “Prostate Pal”, “Get Bladder Fit”, and “UroApp”. Prostate Pal provides an easily accessible tool for recording intake and output, AUA score, and PSA level, but this application showed no validity.

The main issue in mobile research is ensuring the security of patient data. It is an important problem and it has been argued that it is not safe for health care clinicians to have access to patient information from a handheld device [18]. This is the main reason that we designed this Web-based application to ensure that patient information and clinical data were not saved on patients’ own smartphones, but transferred directly to the main server. Administration of the server was only performed by clinicians and, after data were transferred to the hospital server, the records were removed.

The merits of the clinical use of a smartphone application are the possibility of yielding a great number of data points, requiring less time, and more positive compliance by participants than conventional methods, including paper, telephone, and email methods. From a technical view, a smartphone application has computing power, a touch screen, third-party application development and distribution, and high-speed data transfer.

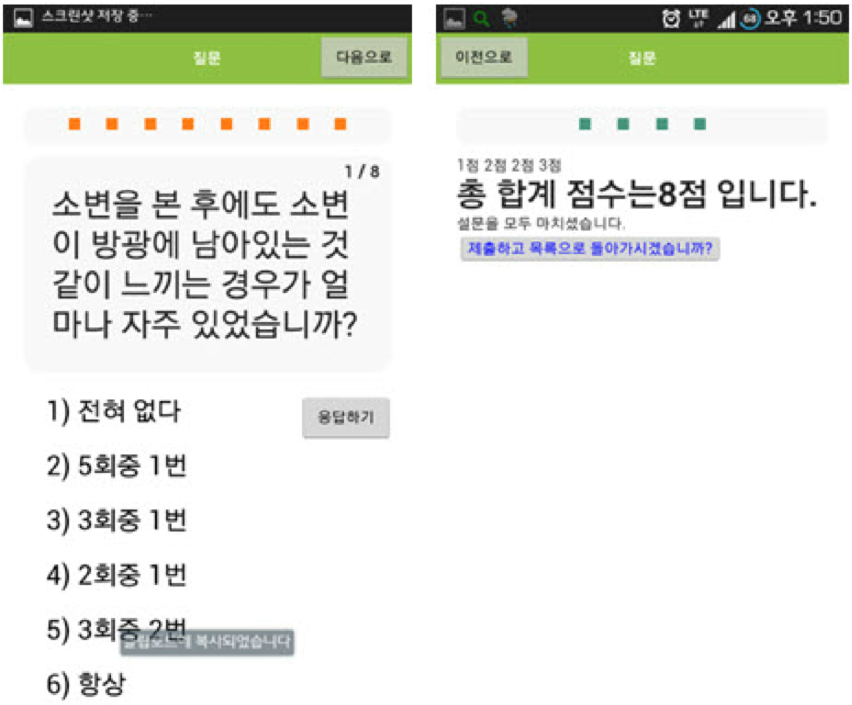
Smartphones can effectively use real-time upload and backup, which can prevent data loss [19,20]. With regard to feasibility, users believed smartphones were easier to use than a conventional paper system in a study of an Android-based application for men’s health [19,21]. In terms of time consumption, a smartphone application has merits over paper questionnaires and Web-based research [12,22,23]. Moreover, increased access and availability to smartphone communication increases the potential for large scale surveys in population-based studies.

We developed this application for the Android operating system. The two main discriminatory factors for determining the operating system were the widespread popularity and homogeneity of the smartphone hardware. We adopted the Android operating system because the number of its users is rapidly growing. With regard to homogenous hardware safety, the iPhone may be a better option.

There are regulatory barriers to the clinical application of smartphone-based medical devices, but smartphone-based questionnaires such as our application are considered Class I devices by the FDA [24,25]. Class I devices represent general devices that are not designed for use in supporting or sustaining life nor are of considerable importance in preventing impairment to human life and have the least demanding restrictions of the three FDA device classes. In recent guidelines, the FDA classified Medical Device Data System (MDDS) software as Class I, because it transfers, stores, converts, or displays medical device data without providing analysis, alarms, or active patient monitoring. This latest set of MDDS guidelines came after the European Commission decision that most applications would be classified under Class I. Our application is a simple substitute for a paper-based questionnaire and does not involve analysis of data, alarms, or active patient monitoring.

Figure 2. Screenshot of the IPSS Android smartphone application. IPSS: International Prostate Symptom Score.

Figure 3. IPSS question 1 and ending screen in the smartphone application. IPSS: International Prostate Symptom Score.



소변을 본 후에도 소변이 방광에 남아있는 것 같이 느끼는 경우가 얼마나 자주 있었습니까?

1) 전혀 없다
2) 5회중 1번
3) 3회중 1번
4) 2회중 1번
5) 3회중 2번
6) 항상

클립보드에 복사되었습니다

1점 2점 2점 3점

총 합계 점수는8점 입니다.

설문을 모두 마치셨습니다.

제출하고 목록으로 돌아가시겠습니까?

Figure 4. Registered IPSS scores are automatically transferred to main server. IPSS: International Prostate Symptom Score.

과민성 방광 증상 점수

번호	질문	09-12
1	아침에 일어나서 밤에 자기 전까지 몇 회 정도 소변을 보셨습니까?	0
2	밤에 잠든 후부터 아침에 일어날 때까지 소변을 보기 위해 몇 회나 일어 나셨습니까?	2
3	갑자기 소변이 마려워 참기 힘들었던 적이 있었습니까?	0
4	갑자기 소변이 마려워서 참지 못하고 소변을 지린 적이 있었습니까?	1
합계		3

국제 전립선 증상 점수

번호	질문	09-12
1	소변을 본 후에도 소변이 방광에 남아있는 것 같이 느끼는 경우가 얼마나 자주 있었습니까?	2
2	소변을 본 후 2시간이내에 다시 소변을 본 경우가 얼마나 자주 있었습니까?	0
3	소변을 볼 때에 소변이 멈추었다가 다시 시작되는 경우가 얼마나 자주 있었습니까?	0
4	소변이 마려울 때 참기 어려운 경우가 얼마나 자주 있었습니까?	0
5	소변 줄기가 약해지거나 가늘어진 경우가 얼마나 자주 있었습니까?	1
6	소변을 볼 때 힘을 줘야 하거나 기다려야 하는 경우가 얼마나 자주 있었습니까?	3
7	밤에 주무시는 동안 소변을 보려고 몇 번이나 잠을 깨십니까?	2

841121	■	상세보기
400114	■	상세보기
680920	■	상세보기
560211	■	상세보기
580222	■	상세보기
320322	■	상세보기
641113	■	상세보기
700820	■	상세보기
630718	■	상세보기
670906	■	상세보기
701114	■	상세보기

Limitations

This study has several limitations. First, there could be selection bias in this study. Some patients reported a mild negative reaction to the questionnaire method. It is possible that smartphone assessment is less suitable in certain subgroups of patients. Some groups, including older-aged people, may use this mobile technology less than others and find it to be a barrier. However, it is estimated that 80-90% of the population will have a smartphone within 10 years [26]. Considering the rapid increase in the use of smartphones, older-aged people will increasingly come to be familiar with smartphones.

There is also skepticism over the possible differences between self-reported measures and clinician-based ratings. The IPSS was originally developed as a self-reported questionnaire and self-report measures may be time and cost-saving methods.

Although we demonstrated the validity of a smartphone application of the IPSS by quantitative analysis, concerns still remain regarding the quality of the analysis, including data recording, data entry, reliability, time consumption, and costs. Our application does not contain open-ended questions, and therefore we did not examine issues of quality, except for compliance and satisfaction rates. A more streamlined graphical and colorful user interface may lead to better compliance for participants for both screening and monitoring.

We divided patients into three groups based on the severity of their symptoms as ascertained by their QOL scores: mild (0-2), moderate (3-4), and severe (5-6). This classification was arbitrary. We did not consider sociodemographic factors of patients that could impact the development of LUTS. Nevertheless, to our knowledge, this is the first study to demonstrate the validity of a smartphone version of the IPSS by scientific methods.

Conclusions

In summary, a smartphone application of the IPSS has validity and reliability, which means that it is not inferior to the paper version of the IPSS. Further research is needed to test its efficacy in the monitoring of LUTS. Future studies are needed to demonstrate its role in consecutive monitoring and also its usefulness in cost savings and data collection.

Acknowledgments

This research was supported by the Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0011678) and Soonchunhyang University Research Fund.

Authors' Contributions

JHK and S-SK were responsible for conception, design, and drafting of the manuscript; HYS, SRS, and D-IC were responsible for acquisition of data; WJY, YMK, WJY, and YSS performed critical revision and supervision; and S-SK conducted the statistical analysis.

Conflicts of Interest

None declared.

References

1. Quek KF, Loh CS, Low WY, Razack AH. Quality of life assessment before and after transurethral resection of the prostate in patients with lower urinary tract symptoms. World J Urol 2001 Nov;19(5):358-364. [Medline: 11760785]
2. Tsang KK, Garraway WM. Impact of benign prostatic hyperplasia on general well-being of men. Prostate 1993;23(1):1-7. [Medline: 7687780]
3. Barry MJ. Evaluation of symptoms and quality of life in men with benign prostatic hyperplasia. Urology 2001 Dec;58(6 Suppl 1):25-32; discussion 32. [Medline: 11750246]
4. Roehrborn CG. Male lower urinary tract symptoms (LUTS) and benign prostatic hyperplasia (BPH). Med Clin North Am 2011 Jan;95(1):87-100. [doi: 10.1016/j.mcna.2010.08.013] [Medline: 21095413]
5. Barry MJ, Cockett AT, Holtgrewe HL, McConnell JD, Sihelnik SA, Winfield HN. Relationship of symptoms of prostatism to commonly used physiological and anatomical measures of the severity of benign prostatic hyperplasia. J Urol 1993 Aug;150(2 Pt 1):351-358. [Medline: 7686980]
6. Barry MJ, Fowler F, O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK. Measuring disease-specific health status in men with benign prostatic hyperplasia. Measurement Committee of The American Urological Association. Med Care 1995 Apr;33(4 Suppl):AS145-AS155. [Medline: 7536866]
7. Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, et al. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. BMC Med Inform Decis Mak 2009;9:51 [FREE Full text] [doi: 10.1186/1472-6947-9-51] [Medline: 20030813]
8. Nielsen. Smartphones to overtake feature phones in US by 2011. 2010. URL: http://www.nielsen.com/us/en/newswire/2010/smartphones-to-overtake-feature-phones-in-u-s-by-2011.html [accessed 2014-01-19] [WebCite Cache ID 6Mk1O1cDN]
9. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
10. Behar J, Roebuck A, Domingos JS, Gederi E, Clifford GD. A review of current sleep screening applications for smartphones. Physiol Meas 2013 Jul;34(7):R29-R46. [doi: 10.1088/0967-3334/34/7/R29] [Medline: 23771214]
11. Palmier-Claus JE, Ainsworth J, Machin M, Barrowclough C, Dunn G, Barkus E, et al. The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application. BMC Psychiatry 2012;12:172 [FREE Full text] [doi: 10.1186/1471-244X-12-172] [Medline: 23075387]
12. Pfaeffli L, Maddison R, Jiang Y, Dalleck L, Löf M. Measuring physical activity in a cardiac rehabilitation population using a smartphone-based questionnaire. J Med Internet Res 2013;15(3):e61 [FREE Full text] [doi: 10.2196/jmir.2419] [Medline: 23524251]
13. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005 Sep;27(3):281-291 [FREE Full text] [doi: 10.1093/pubmed/fdi031] [Medline: 15870099]
14. Choi HR, Chung WS, Shim BS. Translation validity and reliability of I-PSS Korean version. Korean J Urol 1996;37:659-665.
15. Cam K, Akman Y, Cicekci B, Senel F, Erol A. Mode of administration of international prostate symptom score in patients with lower urinary tract symptoms: physician vs self. Prostate Cancer Prostatic Dis 2004;7(1):41-44. [doi: 10.1038/sj.pcan.4500683] [Medline: 14999237]
16. Ainsworth J, Palmier-Claus JE, Machin M, Barrowclough C, Dunn G, Rogers A, et al. A comparison of two delivery modalities of a mobile phone-based assessment for serious mental illness: native smartphone application vs text-messaging only implementations. J Med Internet Res 2013;15(4):e60 [FREE Full text] [doi: 10.2196/jmir.2328] [Medline: 23563184]
17. Hsi RS, Hotaling JM, Hartzler AL, Holt SK, Walsh TJ. Validity and reliability of a smartphone application for the assessment of penile deformity in Peyronie's disease. J Sex Med 2013 Jul;10(7):1867-1873. [doi: 10.1111/jsm.12136] [Medline: 23551808]
18. Haynes CL, Cook GA, Jones MA. Legal and ethical considerations in processing patient-identifiable data without patient consent: lessons learnt from developing a disease register. J Med Ethics 2007 May;33(5):302-307 [FREE Full text] [doi: 10.1136/jme.2006.016907] [Medline: 17470509]
19. Zhang S, Wu Q, van Velthoven MH, Chen L, Car J, Rudan I, et al. Smartphone versus pen-and-paper data collection of infant feeding practices in rural China. J Med Internet Res 2012;14(5):e119 [FREE Full text] [doi: 10.2196/jmir.2183] [Medline: 22989894]

20. Yu P, de Courten M, Pan E, Galea G, Pryor J. The development and evaluation of a PDA-based method for public health surveillance data collection in developing countries. Int J Med Inform 2009 Aug;78(8):532-542. [doi: 10.1016/j.ijmedinf.2009.03.002] [Medline: 19369114]
21. Rajput ZA, Mbugua S, Amadi D, Chepngeno V, Saleem JJ, Anokwa Y, et al. Evaluation of an Android-based mHealth system for population surveillance in developing countries. J Am Med Inform Assoc 2012;19(4):655-659 [FREE Full text] [doi: 10.1136/amiajnl-2011-000476] [Medline: 22366295]
22. Rabin C, Bock B. Desired features of smartphone applications promoting physical activity. Telemed J E Health 2011 Dec;17(10):801-803. [doi: 10.1089/tmj.2011.0055] [Medline: 22010977]
23. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Schizophr Bull 2012 May;38(3):414-425 [FREE Full text] [doi: 10.1093/schbul/sbr155] [Medline: 22080492]
24. US Food and Drug Administration. 2011 Jul 21. USFDA 2011 Draft guidance for industry and food and drug administration staff-mobile medical applications URL: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf [accessed 2014-01-19] [WebCite Cache ID 6Mk2w1yli]
25. US Food and Drug Administration. 2013 Sep 25. USFDA 2013 Is the product a medical device? URL: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/ClassifyYourDevice/ucm051512.htm [accessed 2014-01-19] [WebCite Cache ID 6Mk3i67Xx]
26. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]

Abbreviations

BPH: benign prostatic hyperplasia
ICC: intraclass correlation coefficient
IPSS: International Prostate Symptom Score
LUTS: lower urinary tract symptoms
MDDS: Medical Device Data System

Edited by G Eysenbach; submitted 23.10.13; peer-reviewed by SW Lee, IH Chang, JH Bae; comments to author 15.11.13; revised version received 13.12.13; accepted 23.12.13; published 10.02.14.

Please cite as:
Kim JH, Kwon SS, Shim SR, Sun HY, Ko YM, Chun DI, Yang WJ, Song YS
Validation and Reliability of a Smartphone Application for the International Prostate Symptom Score Questionnaire: A Randomized Repeated Measures Crossover Study
J Med Internet Res 2014;16(2):e38
URL: http://www.jmir.org/2014/2/e38/
doi:10.2196/jmir.3042
PMID:24513507

Original Paper

Applying Computerized Adaptive Testing to the Negative Acts Questionnaire-Revised: Rasch Analysis of Workplace Bullying

Shu-Ching Ma[1,2], RN, MSN; Tsair-Wei Chien[3,4], MBA; Hsiu-Hung Wang[1], RN, PhD, FAAN; Yu-Chi Li[2], RN, MSN; Mei-Shu Yui[5], NP

[1]College of Nursing, Kaohsiung Medical University, Kaohsiung, Taiwan
[2]Nursing Department, Chi-Mei Medical Center, Tainan, Taiwan
[3]Department of Hospital and Health Care Administration, Chia-Nan University of Pharmacy and Science, Tainan, Taiwan
[4]Planning Department, Chi-Mei Medical Center, Tainan, Taiwan
[5]Orthopedics Department, Chi-Mei Medical Center, Tainan, Taiwan

Corresponding Author:
Hsiu-Hung Wang, RN, PhD, FAAN
College of Nursing
Kaohsiung Medical University
No. 100, Shih-Chuan 1st Road
Kaohsiung, 80708
Taiwan
Phone: 886 7 3121101 ext 2624
Fax: 886 7 3218364
Email: hhwang@kmu.edu.tw

Abstract

Background: Workplace bullying is a prevalent problem in contemporary work places that has adverse effects on both the victims of bullying and organizations. With the rapid development of computer technology in recent years, there is an urgent need to prove whether item response theory–based computerized adaptive testing (CAT) can be applied to measure exposure to workplace bullying.

Objective: The purpose of this study was to evaluate the relative efficiency and measurement precision of a CAT-based test for hospital nurses compared to traditional nonadaptive testing (NAT). Under the preliminary conditions of a single domain derived from the scale, a CAT module bullying scale model with polytomously scored items is provided as an example for evaluation purposes.

Methods: A total of 300 nurses were recruited and responded to the 22-item Negative Acts Questionnaire-Revised (NAQ-R). All NAT (or CAT-selected) items were calibrated with the Rasch rating scale model and all respondents were randomly selected for a comparison of the advantages of CAT and NAT in efficiency and precision by paired *t* tests and the area under the receiver operating characteristic curve (AUROC).

Results: The NAQ-R is a unidimensional construct that can be applied to measure exposure to workplace bullying through CAT-based administration. Nursing measures derived from both tests (CAT and NAT) were highly correlated (r=.97) and their measurement precisions were not statistically different (P=.49) as expected. CAT required fewer items than NAT (an efficiency gain of 32%), suggesting a reduced burden for respondents. There were significant differences in work tenure between the 2 groups (bullied and nonbullied) at a cutoff point of 6 years at 1 worksite. An AUROC of 0.75 (95% CI 0.68-0.79) with logits greater than –4.2 (or >30 in summation) was defined as being highly likely bullied in a workplace.

Conclusions: With CAT-based administration of the NAQ-R for nurses, their burden was substantially reduced without compromising measurement precision.

(J Med Internet Res 2014;16(2):e50) doi:10.2196/jmir.2819

KEYWORDS

computerized adaptive testing; computer on wheels; classic test theory; item response theory; nonadaptive testing; the Negative Acts Questionnaire-Revised

Introduction

Background

Workplace bullying is defined as persistent exposure to interpersonal aggression and mistreatment from colleagues, superiors, or subordinates [1,2]. It is a prevalent problem in the workplace with adverse effects on both victims and organizations [3,4]. Many studies have investigated this problem by determining its frequency, identifying groups at risk in different occupational groups and sectors [5], and addressing prevalence of bullying in different countries and among different occupational groups [6]. However, none of these bullied victim evaluations have applied item response theory (IRT) [7] to assess item functioning of the workplace bullying-related questionnaire [8].

Similarly, no studies have reported results on workplace bullying using IRT-based computerized adaptive testing (CAT) to measure respondents’ bullying exposure, especially in the era of computer technology and when questionnaires have become more integrated in recent years. As of April 24, 2013, 127 articles were found on PubMed by searching the keywords “computer adaptive test” (CAT), 309 with “workplace bullying,” and 106 with “workplace bullying nurse”. It is necessary to investigate whether CAT can be applied to yield the same results as traditional nonadaptive testing (NAT) on a workplace bullying scale for nurses and, thereby, reduce their respondent burden.

Computer Assessment and Computer Adaptive Testing

From the literature, traditional paper-and-pencil or computer-based surveys (NAT) have a large respondent burden and require respondents to answer all the questions [9]. In contrast, CAT-based testing using IRT can achieve similar measurement precision levels as NAT and is approximately half the length of the test [10-13]. However, most CAT articles, except some [9,14,15], compared CAT to NAT with dichotomous items. Whether polytomously scored items on bullying can also be measured as precisely as dichotomous CAT should be further investigated.

Rasch Analysis

In classical test theory (CTT), raw scores are usually used as linear interval scale measures for additive latency to assess respondents’ underlying ability. Unfortunately, this is not correct [16,17]; therefore, subsequent statistical analyses can be problematic and incorrect in computing mean, variance, correlation coefficients, or Cronbach alpha [18,19]. In particular, CTT encounters problems when dealing with missing data.

To overcome this obstacle, the IRT-based Rasch model [20] was developed to represent the probabilistic relationship between a person measure and an item difficulty in log-odds units, or logits. A useful scale using the Rasch model should be evaluated by 3 steps (prior tests, Rasch fit statistics, and post hoc tests) suggested by Smith [21] and Tennant and Pallant [22] (details shown in Methods) to verify a single domain. In many articles, authors used Rasch modeling to develop CAT on clinical samples, but none adopted the model testing steps recommended by Smith to verify scales before implementing CAT [9,10,23-26].

Objectives

First, we used a polytomous Rasch rating scale model to examine the workplace bullying scale for CAT use. Second, we developed an Excel Visual Basic for Applications (VBA) CAT module for comparison with NAT on efficiency, precision, and inference from the data of 300 hospital nurses. Third, similar inferences made by CAT and NAT were conducted in addition to investigating significant differences in work tenure between 2 groups. Fourth, a cutoff point of the studied bullying scale was determined for discriminating persons who were bullied victims with a predicted (individual) probability.

We report the CAT advantages if the precision and inference of results made by CAT and NAT are similar. Several limitations of CAT application will be raised for consideration in future studies.

Methods

Study Participants

The study sample was randomly selected and recruited using the last 3 digits of the identification card number from nurses of a 1333-bed medical center in southern Taiwan in the summer of 2010. No incentive for participation was offered. A total of 300 nurses completed 2 effective eligibility scales (shown in the following section) using NAT. This study was approved and monitored by the Research Ethics Review Board of the Chi-Mei Medical Center.

Demographic data collected included gender, work tenure in hospitals of all types, age, marital status, and education level.

Scales Used

Exposure to Bullying

The Negative Acts Questionnaire-Revised (NAQ-R) [27] used in this study has 22 items with 5 response alternatives (1=never, 2=occasionally, 3=monthly, 4=weekly, 5=daily) to measure exposure to workplace bullying within the past 6 months. Victimization from bullying during the past 6 months was additionally measured by a single self-labeling victimization question that was used for determining the cutoff point of the studied bullying scale after bullying measures were obtained. The NAQ-R was professionally translated into Chinese by authors in Taiwan using a back-translation technique (English-Chinese-English). With permission from the author [28], we conducted Rasch analysis to test scale unidimensionality (shown in the dimensionality section), the appropriateness level of the 5-category NAQ-R [29], as well as reporting reliability (Cronbach alpha) and dimension coefficient (DC) [30] using the CTT method.

Negative Actions Caused by Bullying

Participants were asked questions about their own personal negative experience of bullying and its impact on 5 areas (physical aspects, psychological aspects, interpersonal relations at work, willingness to work, and quality of work) and they were asked to respond to personal symptoms or emotions (eg, gastrointestinal symptoms, fatigue, loss of appetite, crying, fear, anxiety, no sense of belonging, absenteeism, intent to leave the job, hating work, not being able to concentrate on work, loss of patience when caring for patients, frequent occurrence of abnormalities, low self-esteem, sleep disorders, anxiety, concentration disorders, chronic fatigue, anger, depression, several somatic disorders) [31], all of which were subjective self-judgments (yes=1; no=0) and were evolved into a global scale to verify discriminant validity of the NAQ-R.

Dimensionality

Tennant and Pallant [22] reported that 3 steps should be applied to assess scale unidimensionality: (1) conduct prior testing using Horn’s parallel analysis [32] to make sure that a single dimension is suitable; (2) use Rasch fit statistics ranging from 0.5 to 1.5 [33,34] to determine the usefulness of the 1-dimensional scaling; and (3) run post hoc tests using the first principal components analysis (PCA) component of Rasch standardized residuals [35] close to zero to inspect the convergent validity, and then performing Smith [36] independent *t* tests to compare estimates of the percentages (<5%, within ±1.96) and verify invariance of Rasch model (details presented in following section).

Differential Item Functioning

The Rasch rating scale model (used in this study) requires the item estimation to be independent of the subgroups of individuals completing the questions. In other words, item parameters should be invariant across populations [37]. Items not demonstrating invariance are commonly referred to as exhibiting differential item functioning or item bias.

The chi-square test used for detecting the item-trait interaction was computed from a comparison of the observed overall performance of each trait group on the item with its expected performance [38]. Its probability (eg, <.05) reports the statistical probability of observing the chi-square value (or worse) when the data fit the Rasch model. Thus, WINSTEPS table 3.4 was referred to detect differential item functioning items for a significantly different group of person measures [39].

Computer Adaptive Test Procedures and Features

We ran a VBA module in Microsoft Excel in compliance with rules and regulations of CAT (Figures 1 and 2). Cronbach alpha and Rasch person separation reliability calculated from the NAQ-R of the study were used to determine the CAT termination criterion using the standard error of measurement (SEM=SD × √reliability), whereas Rasch reliability refers to reliability in the previously mentioned SEM formula.

We also set another rule that the minimum number of questions required for completion was 10 (10/22 items on NAQ-R item length=45%) because CAT could achieve similar precision in measurement as NAT with approximately half the length [9-12]. The first question was selected randomly from the 22 items when performing the CAT. The provisional measures were estimated by a maximum log likelihood function using an iterative Newton-Raphson procedure [9,12] after 3 questions were answered without responding similarly and sequentially to either 1 or 5. The next question selected was the one with the most information obtained from the remaining unanswered questions, interacting with the provisional person measures. All responses and their respective consumption time for each nurse were recorded after CAT termination.

Figure 1. Computer adaptive test applied to Negative Acts Questionnaire-Revised (NAQ-R) in workplace.

CAT based Negative Acts Questionnaire-Revised scale

No	Measure	response	|Z-score|
5	-1.17	2	1.03
9	-0.79	3	2.02
6	-1.02	2	0.13
4	-0.99	1	2.09
2	-0.59	2	0.97
3	-0.28	3	0.88
10	-0.33	2	0.47
1	0.17	3	0.6
22	0.08	2	1.06
16	0	3	1.09

Item	Items	difficult
1	Someone withhol	-2.21
2	Being humiliated c	-0.79
3	Being ordered to	-0.24
4	Having key areas	-0.09
5	Spreading of goss	-0.9
6	Being ignored or e	0.61
7	Having insulting o	-0.19
8	Being shouted at	-1.5
9	Intimidating beha	1.35
10	Hints or signals frc	1.65
11	Repeated reminde	-0.68
12	Being ignored or f	-0.57
13	Persistent criticism	-0.33
14	Having your opini	-0.61
15	Practical jokes car	0.46
16	Being given tasks	0.04
17	Having allegations	0.2
18	Excessive monito	-0.41
19	Pressure not to cl	0.37
20	Being the subject	1.09
21	Being exposed to	0.18
22	Threats of violenc	2.56

1.NO: 000010 + Date: 2013/04/23

email:

2.Over views on Bullying

○ Never ○ Very Few ○ Now and then ● Monthly ○ Weekly' ○ Daily ○ Always

Spreading of gossip and rumours about you

Measure -1.17

SE 0.44

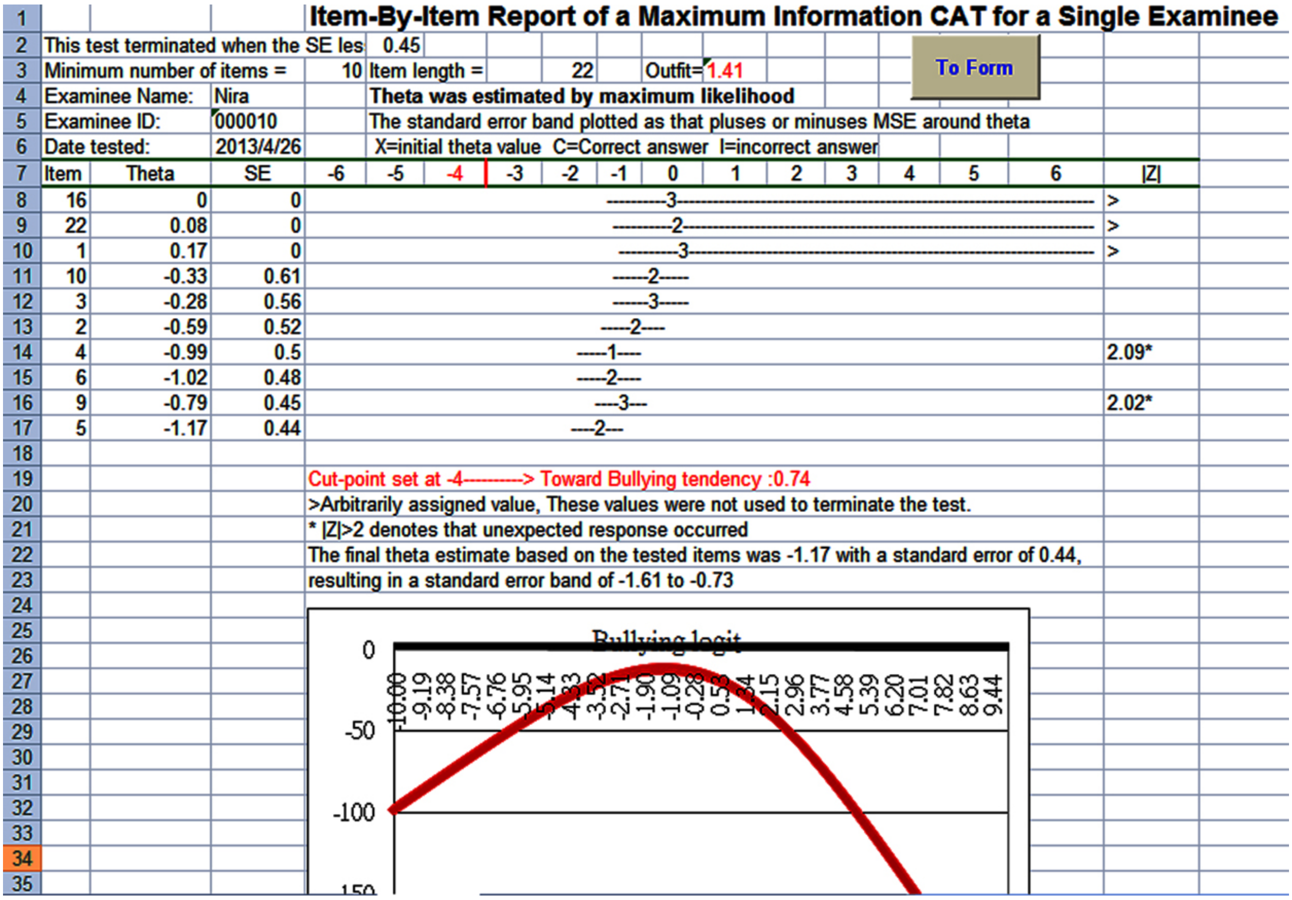
Inform 5.17

Outfit 1.41

Infit 1.36

Item 10

Figure 2. Bullying report produced by Negative Acts Questionnaire-Revised (NAQ-R) computer adaptive test with a maximum likelihood estimation plot.



Item-By-Item Report of a Maximum Information CAT for a Single Examinee

This test terminated when the SE les 0.45

Minimum number of items = 10 Item length = 22 Outfit=1.41

Examinee Name: Nira

Theta was estimated by maximum likelihood

Examinee ID: 000010

The standard error band plotted as that pluses or minuses MSE around theta

Date tested: 2013/4/26

X=initial theta value C=Correct answer I=incorrect answer

Item	Theta	SE	-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6	|Z|
16	0	0	---------3---	>
22	0.08	0	---------2--	>
1	0.17	0	---------3--	>
10	-0.33	0.61	------2-----	
3	-0.28	0.56	------3-----	
2	-0.59	0.52	-----2----	
4	-0.99	0.5	-----1----	2.09*
6	-1.02	0.48	-----2----	
9	-0.79	0.45	----3---	2.02*
5	-1.17	0.44	----2---	

Cut-point set at -4---------> Toward Bullying tendency :0.74

>Arbitrarily assigned value, These values were not used to terminate the test.

* |Z|>2 denotes that unexpected response occurred

The final theta estimate based on the tested items was -1.17 with a standard error of 0.44, resulting in a standard error band of -1.61 to -0.73

Comparison of Efficiency and Precision

Four indexes between CAT and NAT were compared, including test length (efficiency), estimated measures (precision), time saved (in seconds) per item (efficiency), and the area under the receiver operating characteristic (AUROC) curve (precision).

Accordingly, all person measures based on NAT should be estimated in advance, assuming all 22 items were answered. The following steps were adopted: (1) using the WINSTEPS software [39] to calibrate item and threshold difficulties, and (2) performing the studied dataset of 300 people × 22 items to re-estimate both NAT (through all 22 items) and CAT (by less than 22 items) measures using the CAT Excel-VBA module.

Comparison of Groups by Making Inferences

We compared the prediction effects of CAT and NAT on the 4 indexes by regressing person measures, respectively, on (1) the global symptom score, and (2) differences in demographic characteristics (eg, age, work tenure, and marital status) and in self-judgments (victimization from bullying during the past 6 months).

Statistical Analysis

For all statistical analyses, SPSS software for Windows version 12 (SPSS, Chicago, IL, USA) was used. The CAT and NAT person measures were compared using the Pearson correlation coefficient. Test length (efficiency) and estimated measures (precision) were compared by paired *t* tests. Time saved per item (efficiency) in favor of CAT was computed by a margin of 25.07 seconds (SD 16.04; range 12-43). The total time saving from NAT to CAT was computed by the formula: time saved per item (25.07) multiplies both item lengths shortened by CAT (eg, 2109 items in total), and the sample size (N=300).

The AUROC (precision) was calculated by both Rasch-transformed logit scores and the single self-labeled victimization question from bullying (bullied=1; not bullied=0) to determine a cutoff point with 95% confidence intervals (CIs), sensitivity, and specificity. The criterion of alpha=.05 was considered statistically significant. Horn’s parallel analysis was performed using an online calculator [40] that is based on the literature [32,41].

Results

Overview

Two age groups (separated by a cutoff point of 30 years) were contrasted on demographic characteristics (eg, self-labeled victimization from bullying, gender, work tenure within the study hospital, work tenure in health care, marital status, and education level). As seen in Table 1, the prevalence of bullying within the study hospital was 24.0% (72/300). CAT and NAT measures were highly correlated (*r*=.97)

Table 1. Demographic characteristics of participants (N=300).

Characteristics	Age (years), n (%)		Total, n
	<30	≥30	
Age	109 (36.3)	191 (63.7)	300
Bullied or not			
No	84 (36.8)	144 (63.2)	228
Yes	25 (34.7)	47 (65.3)	72
Gender			
Male	1 (25.0)	3 (75.0)	4
Female	108 (36.5)	188 (63.5)	296
Work tenure in the hospital (years)			
≤3	50 (92.6)	4 (7.4)	54
3-6	47 (69.1)	21 (30.9)	68
6-9	11 (16.2)	57 (83.8)	68
>9	1 (0.9)	109 (99.1)	110
Work tenure in health care (years)			
≤3	46 (95.8)	2 (4.2)	48
3-6	47 (83.9)	9 (16.1)	56
6-9	15 (21.7)	54 (78.3)	69
>9	1 (0.8)	126 (99.2)	127
Marital status			
Married	22 (16.7)	110 (83.3)	132
Unmarried	87 (51.8)	81 (48.2)	168
Education diploma			
Undergraduate	109 (37.7)	180 (62.3)	189
Postgraduate	—	11 (100)	11

Dimensionality

The NAQ-R can be considered unidimensional given that (1) one factor was extracted by parallel analysis; (2) all infit and outfit mean squares for the 22 items were in a range of 0.5 to 1.5 (shown in Table 2); (3) item loadings from the Rasch PCA of residuals on the first contrast were closely clustered within 0.6 or near 0.6; PTME (ie, point measure regarding the Pearson correlation between the observations of an item and the item difficulties that is similar to factor loading) between 0.48 and 0.78; Rasch person separation reliability=0.90, Cronbach alpha=.98 (>.70), DC=0.89 (>0.70), and Smith's *t* test of proportions [36] reach zero outside the range ±1.96 (ie, all persons' paired *t* values were within ±1.96). In addition, category structure for the NAQ-R should display the monotonically increasing threshold (-3.39, -0.55, 1.11, and 2.83 logits; Figure 3) with the Linacre's guidelines [29] to improve the utility of the resulting measures. The absence of differential item functioning suggests good support for measurement invariance. The range of threshold difficulties for those least difficult items 1 and 8 are shown as examples in 2 columns on the right-hand side in Figure 4, indicating that item difficulties cannot sufficiently cover all the person measures with mild or nonbullied symptoms shown on the left bottom in Figure 4 using the NAQ-R scale.

Table 2. One factor extracted from the Negative Acts Questionnaire-Revised (NAQ-R) scale with mean square between 0.50 and 1.50.

During the last 6 months, how often have you been subjected to the following negative acts in the work place?	Item		Mean square		Rasch
	Delta	PTME	Infit	Outfit	Loading
Work-related bullying					
19. Pressure not to claim something to which by right you are entitled	0.37	0.68	1.22	0.88	–0.26
21. Being exposed to an unmanageable workload	0.18	0.67	1.22	1.12	–0.27
16. Being given tasks with unreasonable deadlines	0.04	0.71	0.94	0.95	–0.21
3. Being ordered to do work below your level of competence	–0.24	0.68	1.11	1.10	0.59
18. Excessive monitoring of your work	–0.41	0.77	0.93	0.84	–0.43
14. Having your opinions ignored	–0.61	0.75	0.87	0.92	–0.45
1. Someone withholding information which affects your performance	–2.21	0.73	1.21	1.28	0.47
Person-related bullying					
10. Hints or signals from others that you should quit your job	1.65	0.60	0.96	0.76	0.03
20. Being the subject of excessive teasing and sarcasm	1.09	0.69	0.75	0.63	–0.30
6. Being ignored or excluded	0.61	0.68	0.93	0.90	0.17
15. Practical jokes carried out by people you don’t get along with	0.46	0.72	0.81	0.65	–0.32
17. Having allegations made against you	0.20	0.73	0.81	0.67	–0.32
4. Having key areas of responsibility removed or replaced with more trivial or unpleasant tasks	–0.09	0.69	1.12	1.08	0.31
7. Having insulting or offensive remarks made about your person, attitudes or your private life	–0.19	0.72	0.99	0.90	0.14
13. Persistent criticism of your errors or mistakes	–0.33	0.78	0.77	0.65	–0.49
12. Being ignored or facing a hostile reaction when you approach	–0.57	0.72	1.17	1.04	–0.34
11. Repeated reminders of your errors or mistakes	–0.68	0.76	0.95	0.90	–0.26
2. Being humiliated or ridiculed in connection with your work	–0.79	0.72	1.07	0.98	0.59
5. Spreading of gossip and rumors about you	–0.90	0.75	0.97	0.96	0.31
Physically intimidating bullying					
22. Threats of violence or physical abuse or actual abuse	2.56	0.48	1.49	0.56	–0.12
9. Intimidating behaviors such as finger-pointing, invasion of personal space, shoving, blocking your way	1.35	0.61	1.06	0.97	0.11
8. Being shouted at or being the target of spontaneous anger	–1.50	0.71	1.37	1.36	0.25
Minimum	–2.21	0.48	0.75	0.56	–0.49
Maximum	2.56	0.78	1.49	1.36	0.59

Figure 3. Threshold step difficulties monotonically increasing for the Negative Acts Questionnaire-Revised (NAQ-R).

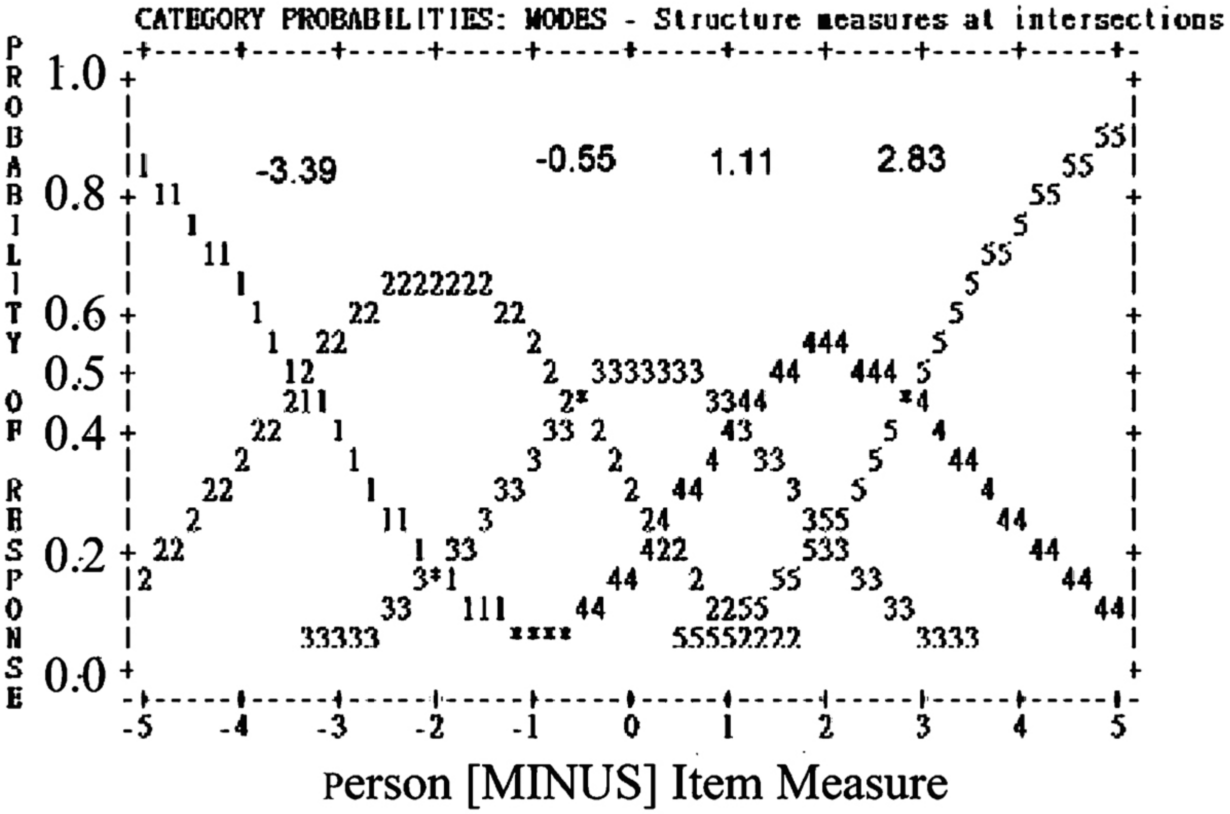


Figure 4. Variable map for person and item dispersion on Rasch logit scores.

Comparison of Efficiency and Precision

The CAT required substantially fewer items than NAT (*P*<.001). NAT required all 300 participants to respond to all 22 items, yielding 6600 responses. In CAT, only 4491 responses were required, meaning that each nurse answered 14.97 questions on average. As compared to NAT, CAT provided an efficient gain in test length of 0.32, calculated as 1–(ratio of total responses by CAT and NAT), or 1–(4491/6600).

For precision of measurement, person measures from CAT did not statistically differ from those from NAT (*P*=.14). The total time saving from NAT to CAT was 52,848 seconds (25.07×7.03×300) or 14.68 hours.

Cutoff point values were >–4.2 logits (approximately >30 in traditional summation); AUROC (95% CI), sensitivity, and specificity were found to be similar between CAT and NAT.

Similar Results in Making Inferences

Rasch logit measures (*x*) of CAT and NAT can yield similar response slope parameters to predict the global scores (*y*) for negative actions. A 2-way ANOVA revealed that person measures only differed in groups of the bullied and the nonbullied as well as groups with job tenure of less than and more than 30 years (in the study hospital) (Table 3).

Table 3. Similar inferences made by computer adaptive testing (CAT) and nonadaptive testing (NAT).

Demographic characteristics	NAT			CAT		
	Mean square	$F_{1,1}$	P	Mean square	$F_{1,1}$	P
Age						
Age	0.76	0.13	.72	1.21	0.21	.65
Victim	265.60	46.73	<.001	326.19	55.14	<.001
Age × victim	10.53	1.85	.18	21.87	3.69	.06
Residual	5.68			5.92		
Bullied in logit						
Yes	–3.09	—		–2.86	—	
No	—	–5.41		—	–5.42	
Work tenure 1 (in this hospital)						
Tenure	23.59	4.18	.11	30.96	5.26	.02
Victim	267.53	47.41	<.001	327.25	55.56	<.001
Tenure × victim	6.53	1.16	.28	14.99	2.55	.11
Residual	5.64			5.89		
Tenure (years) in logit						
<6	–3.68	—		–3.68	—	
≥6	—	–4.46		—	–4.46	
Work tenure 2 (in health care)						
Tenure	16.22	2.86	.09	22.52	3.81	.05
Victim	243.25	42.94	<.001	307.68	51.96	<.001
Tenure × victim	2.75	0.49	.49	9.95	1.68	.19
Residual	5.67			5.92		
Marital status						
Marry	0.29	0.05	.82	0.05	0.01	.93
Victim	259.03	45.31	<.001	306.54	51.15	<.001
Marry × victim	0.57	0.10	.75	0.44	0.07	.79
Residual	5.72			5.99		

Discussion

Key Findings

The results from this study indicate that the 22-item NAQ-R is considered unidimensional. The CAT is up to 32% more efficient for answering questions and achieved similar precision and inferences in measurements as did NAT. A cut point of >–4.2 logits (or >30 in summation) with AUROC 0.75 (95% CI 0.68-0.79) was determined for future use in workplace bullying surveys. The prevalence of bullying for the study sample was 0.24.

What This Adds to What Was Known

Consistent with the literature [9-12], the efficiency of CAT over NAT was supported. We confirm the CAT-based NAQ-R requires significantly fewer questions to measure victimization from workplace bullying than NAT without compromising its measurement precision.

What It Implies and What Should Be Changed?

Easy to Detect Unexpected Responses

The CAT module can help us efficiently and precisely gather responses from nurses and it was technically applicable. Outfit mean square values of 2.0 or greater can be used to examine whether responses are distorted or abnormal. That is, much more unexpected responses deemed to occur because of possibly careless, mistaken, or awkward endorsement were found in the measurement [9,10,29] (eg, nurse A gained outfit 1.41 and gave unexpected responses on items 4 and 9 as shown in Figure 2). It is easier to detect problematic responses by using CAT than CTT. Multimedia Appendix 1 is a CAT module that can be downloaded and practiced by interested readers.

Steps to Detect Scale Dimensionality Used for Computer Adaptive Testing

Some studies [2,3,5,8] reported that there were 2 or 3 factors extracted from the NAQ-R because it used the eigenvalue greater than 1 (K_1) rule to extract a number of factors. A number of studies have shown that the K_1 rule is inaccurate and tends to overfactor [32,42,43]. In contrast, Zwick and Velicer's [44] comparison concluded that parallel analysis was the most accurate evaluation method and it was correct 92% of the time (greater than 22% using K_1). That explained why the factor determined using the parallel analysis method in the present study was different from others.

We also uniquely examined it using Smith's [21] recommended other 2 steps (Rasch analyses shown in Methods) to detect scale dimensionality. Compared to the traditional way of using either parallel analysis or Kayser's rule to detect the number of factors, Rasch-based analysis is superior in studying the dimensionality of a given instrument (eg, infit and outfit criteria and PCA residuals). Accordingly, the CAT module can be implemented after the scale unidimensionality and item difficulties have been determined using the Smith's 3 steps and Rasch analysis.

Cutoff Point Recommended for Determining Bully Victims

The AUROCs (0.74 and 0.75 for NAT and CAT, respectively) were not as high as expected (>0.80). It might be acceptable in social science when it is greater than 0.70 because the single self-labeled victimization question (bullied=1; not bullied=0) might be subjectively answered with some bias by examinees' personal perception of bullying.

Regarding another issue that the cutoff point of –4.2 logits is too low to be confident in the lower 24% prevalence of bullying for the study sample, we can see the visualized person and item map in Figure 4. The person sample is not dispersed as normal (with mean 0 and SD 1) as we expected, so that most nurses earn low Rasch-transformed scores. It is because items on the top right-hand side are presented as difficult for nurses to respond to.

In addition, we can use individual Rasch-transformed logit scores to predict their probability of the bullied victimization using the formula: probability=exp(theta–delta)/(1+exp(theta–delta)), where theta=person measure and delta=item difficulty at cutoff point. For instance, a person with –1.5 logits in bully measurement will present his/her probability at 0.94, calculated as exp(–1.5–(–4.2))/(1+exp(–1.5–(–4.2))), where the item difficulty at cutoff point=–4.2 and the specified person measure=–1.5). The 95% confidence intervals are determined by the dispersion of the person's measured error (ie, the value of 1.96×SE transformed to the previously mentioned probability formula).

Strengths of This Study

Using Computer Adaptive Testing to Endorse the NAQ-R

There are 2 major types of standardized assessments in clinical settings [45]: (1) a lengthy questionnaire that requires significant amounts of time and training for administration to achieve a precise assessment, and (2) a rapid, short-form one that briefly screens for the most common symptoms using cutoff points to determine degrees of impairment [46,47]. CAT has the advantages of both types: precision and efficiency. This paper used the Rasch rating scale model (instead of dichotomy or Rasch partial credit model) to design CAT and then applied it to endorse the NAQ-R. We conducted an actual CAT survey procedure (see the module in Multimedia Appendix 1) instead of CAT simulation as other published studies.

Detecting the Appropriateness of Level of Scaling

If the item threshold difficulties (calibrated by Rasch rating scale model) collapsed, categories should be combined to be more efficient for respondents to answer [29]. Unlike NAQ-R on which the responses "about weekly" and "about daily" were subjectively thrown together into one category [48], this study used the Rasch model for detecting the appropriateness of level of scaling [49].

Unique Features

Although the efficiency of a CAT has been well validated in the literature, the findings of this study do not appear to contribute any important information on the CAT approach. In this study, 2 unique features were reported to readers: (1) the 22-item polytomous NAQ-R is suited for CAT administration in practice, and (2) the module of animation CAT presented in Multimedia Appendix 1 is available for interested readers to practice, which is rare in any previously published articles.

Limitations of the Study

Issues for Further Consideration in Future Research

Several issues should be considered more thoroughly in further studies. First, few male nurses were included in the sample so that the differential item functioning for gender could not be identified by Rasch analysis. Second, there is potential sampling bias in this study. More studies are needed to assess the generalizability of the study with different samples and in different institutes using the Chinese version of NAQ-R. Third, the prevalence of bullying in this study hospital was 24%, higher than seen in studies of Japanese nurses (19%) [8], Italian employees (15.2%) [50], and general service workers in general services (from 2% to 17%) [51]. Fourth, more objective estimates of the prevalence for bullying based on the Leymann criterion [52] is worthy of carrying out in future because the self-labeling approach in this study might produce some biases [50].

Computer Adaptive Testing Stopping Rules Used in This Study

The CAT stopping rules are usually determined by SEM and/or no more than a specific number of items needed for achieving both precision and fast assessment. We applied the former and set minimal items for an acceptable level of person conditional reliability in CAT results.

In Figure 2, we demonstrated a CAT example terminated at SE <0.44, calculated as $\sqrt{(1–0.80)}$, where reliability is set at 0.80, instead of 0.32, calculated as $\sqrt{(1–0.90)}$, where reliability is set at 0.90. If using the latter criterion of 0.32, the item length in CAT will approach the total of 22 items. One way to improve the CAT efficiency and precision is to add more easy bullying questions to the scale (see Figure 4), especially for item difficulties located around the cutoff of –4.2 logits to increase the power of diagnostic discrimination for the bullied victims.

Another way is to temporarily lower the acceptable level of precision to 0.80 reliability as was done in this study.

Lenient Criterion to Support a Good Model-Data-Fit

According to the literature [34], the range of 0.6 and 1.4 is recommended for rating scales (Likert/survey). Item 22 has a slightly high value of infit mean square error (mean square error=1.49) which is <1.5, but a lower outfit mean square error of 0.56. The high value of infit mean square error suggests that those nurses with highly negative actions caused by bullying have a sensitive misfit to item 22, but will not be influenced by the too-low cutoff score at -4.2 logits. In contrast, the low outfit mean square error shown in Table 2 indicates that item 22 has a good model-data-fit for most general nurses. WINSTEPS' guidelines [33] supports that a mean square error >1.5 suggests a deviation from unidimensionality in the data. The other 2 model testing steps Smith recommended also verified that the NAQ-R 22-item scale is unidimensional and suggests that it suits CAT administration.

Applications

Developing an Online Computer Adaptive Testing System

Many issues should be further explored in the future, including studies addressing the limitations noted previously and subsequently. For example, the CAT module should be extended to the Internet for easy use so that the NAQ-R can be administered in more workplaces.

Applying the Animation Computer Adaptive Testing Module

One of the important advantages of CAT scoring is that the item pool for the 22-item NAQ-R can be expanded to match a wide range of participants covering different kinds of bullied workers without changing the module and measurement accuracy. The CAT users may also expand the NAQ-R item pools or replace them with other kinds of workplace bullying scales. It must be noted that (1) overall (ie, on average) and step (threshold) difficulties of the questionnaire must be calibrated in advance using Rasch analysis, (2) pictures and audio files for each question shown in the CAT Excel-module should be well-prepared and put in an appropriate folder that can be shown simultaneously to correspond to questions for the animation CAT module, and (3) pictures and audio files included in the CAT need to match original meaning of the items as much as possible to avoid distorting the validity of the scale.

If readers would like to conduct Rasch partial credit model for the NAQ-R, the distinct threshold step difficulties across items should be reset in specific Excel cells in Multimedia Appendix 1.

More Items Added to Decrease Standard Error

We described the category structure in Figure 3 displaying the NAQ-R monotonically increasing threshold (–3.39, –0.55, 1.11, and 2.83 logits). The Rasch rating scale model indicates each item has a common threshold difficulty. Therefore, the overall difficulties (ie, delta in Table 2) for each item plus the threshold step difficulties (eg, items 1 and 8 in Figure 4) form its own range of item difficulties, and only items 5, 8, and 1 with difficulty ranges include the cutoff point at –4.2 logits. To decrease the person's measured error (ie, SE), we suggest the NAQ-R 22-item scale should add more easy items in the future to increase individual person conditional reliability.

Conclusion

The CAT-based NAQ-R forming a unidimensional construct reduces respondents' burden without compromising measurement precision and increases endorsement efficiency. The computer module developed by the authors is recommended for assessing workers with scores beyond a cut point of >–4.2 logits (or >30 in summed score), who should be treated with more concern as soon as possible at an earlier stage.

Acknowledgments

We would like to thank the study participants for their contributions and the participating hospital's support in facilitating this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Excel VBA CAT module offered to interested readers for own practice (bullying.zip).

[ZIP File (Zip Archive), 2MB - jmir_v16i2e50_app1.zip]

References

1. Einarsen S, Hoel H, Notelaers G. Measuring exposure to bullying and harassment at work: Validity, factor structure and psychometric properties of the Negative Acts Questionnaire-Revised. Work & Stress 2009 Jan;23(1):24-44. [doi: 10.1080/02678370902815673]
2. Leymann H. The content and development of mobbing at work. European Journal of Work and Organizational Psychology 1996 Jun;5(2):165-184. [doi: 10.1080/13594329608414853]

3. Einarsen S. Bullying and Emotional Abuse in the Workplace: International Perspectives in Research and Practice. New York: Taylor & Francis; 2003.
4. Andersson LM, Aquino K, Barling J, Baron RJ, Bennett R, Bies RJ, et al. Bullying at work: a perspective from BritainNorth America. In: Fox S, Spector PE, editors. Counterproductive Work Behavior: Investigations Of Actors And Targets. Washington, DC: American Psychological Association (APA); 2005.
5. Zapf D, Einarsen S, Hoel H, Vartia M. Empirical findings on bullying in the workplace. In: Einarsen S , Hoel H , Zapf D , Cooper CL , editors. Bullying and Emotional Abuse in the Workplace: International Perspectives in Research and Practice. New York: Taylor & Francis; 2003.
6. Mikkelsen EG, Einarsen S. Bullying in Danish work-life: Prevalence and health correlates. European Journal of Work and Organizational Psychology 2001 Dec;10(4):393-413. [doi: 10.1080/13594320143000816]
7. Lord FM. Applications of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: L Erlbaum Associates; 1980.
8. Abe K, Henly SJ. Bullying (ijime) among Japanese hospital nurses: modeling responses to the revised Negative Acts Questionnaire. Nurs Res 2010;59(2):110-118. [doi: 10.1097/NNR.0b013e3181d1a709] [Medline: 20216013]
9. Chien TW, Wang WC, Huang SY, Lai WP, Chow JC. A web-based computerized adaptive testing (CAT) to assess patient perception in hospitalization. J Med Internet Res 2011 Aug;13(3):e61 [FREE Full text] [doi: 10.2196/jmir.1785] [Medline: 21844001]
10. Chien TW, Wu HM, Wang WC, Castillo RV, Chou W. Reduction in patient burdens with graphical computerized adaptive testing on the ADL scale: tool development and simulation. Health Qual Life Outcomes 2009;7:39 [FREE Full text] [doi: 10.1186/1477-7525-7-39] [Medline: 19416521]
11. Wainer HW, Dorans NJ. Computerized Adaptive Testing: A Primer. Hillsdale, NJ: L Erlbaum Associates; 1990.
12. Embretson S, Reise S, Reise SP. Item Response Theory for Psychologists. Mahwah, NJ: L Erlbaum Associates; 2000.
13. Weiss DJ. Improving measurement quality and efficiency with adaptive testing. Applied Psychological Measurement 1982 Sep 01;6(4):473-492. [doi: 10.1177/014662168200600408]
14. Hsueh IP, Chen JH, Wang CH, Hou WH, Hsieh CL. Development of a computerized adaptive test for assessing activities of daily living in outpatients with stroke. Phys Ther 2013 May;93(5):681-693. [doi: 10.2522/ptj.20120173] [Medline: 23329557]
15. Hsueh IP, Chen JH, Wang CH, Chen CT, Sheu CF, Wang WC, et al. Development of a computerized adaptive test for assessing balance function in patients with stroke. Phys Ther 2010 Sep;90(9):1336-1344 [FREE Full text] [doi: 10.2522/ptj.20090395] [Medline: 20592269]
16. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. Arch Phys Med Rehabil 1989 Nov;70(12):857-860. [Medline: 2818162]
17. Wang WC. Recent Developments in Rasch Measurement. Hong Kong: The Hong Kong Institute of Education Press; 2010.
18. Crocker L, Algina J. Introduction to Classical and Modern Test Theory. New York: Holt, Rinehart, and Winston; 1986.
19. Nunnally JC, Bernstein IH. Psychometric Theory. New York: McGraw-Hill; 1994.
20. Rasch G. Rasch Measurement Transactions, Part 1. Chicago: Univ of Chicago Social Research; 1993.
21. Smith RM. A comparison of methods for determining dimensionality in Rasch measurement. Structural Equation Modeling: A Multidisciplinary Journal 1996 Jan;3(1):25-40. [doi: 10.1080/10705519609540027]
22. Tennant A, Pallant JF. Unidimensionality matters! (A tale of two Smiths?). Rasch Meas Trans 2006;20(1):1048-1051 [FREE Full text]
23. Velozo CA, Wang Y, Lehman L, Wang JH. Utilizing Rasch measurement models to develop a computer adaptive self-report of walking, climbing, and running. Disabli Rehabil 2008;30(6):458-467. [doi: 10.1080/09638280701617317.787634494]
24. Lehman LA, Woodbury M, Shechtman O, Wang YC, Pomeranz J, Gray DB, et al. Development of an item bank for a computerised adaptive test of upper-extremity function. Disabil Rehabil 2011 Oct;33(21-22):2092-2104. [doi: 10.3109/09638288.2011.560336] [Medline: 21401332]
25. Öztuna D, Elhan AH, Küçükdeveci AA, Kutlay S, Tennant A. An application of computerised adaptive testing for measuring health status in patients with knee osteoarthritis. Disabil Rehabil 2010 Jan;32(23):1928-1938. [doi: 10.3109/09638281003777572] [Medline: 20384449]
26. Elhan AH, Oztuna D, Kutlay S, Küçükdeveci AA, Tennant A. An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. BMC Musculoskelet Disord 2008;9:166 [FREE Full text] [doi: 10.1186/1471-2474-9-166] [Medline: 19094219]
27. Nielsen MB, Skogstad A, Matthiesen SB, Glasø L, Aasland MS, Notelaers G, et al. Prevalence of workplace bullying in Norway: Comparisons across time and estimation methods. European Journal of Work and Organizational Psychology 2009 Mar;18(1):81-101. [doi: 10.1080/13594320801969707]
28. Einarsen S, Skogstad A. Bullying at work: Epidemiological findings in public and private organizations. European Journal of Work and Organizational Psychology 1996 Jun;5(2):185-201. [doi: 10.1080/13594329608414854]
29. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas 2002;3(1):85-106. [Medline: 11997586]
30. Chien TW. Cronbach's alpha with the dimension coefficient to jointly assess a scale's quality. Rasch Meas Trans 2012;26(3):1379 [FREE Full text]

31. Einarsen S, Raknes BI, Matthiesen SB. Bullying and harassment at work and their relationships to work environment quality: An exploratory study. European Work and Organizational Psychologist 1994 Oct;4(4):381-401. [doi: 10.1080/13594329408410497]
32. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965 Jun;30(2):179-185. [doi: 10.1007/BF02289447]
33. Linacre JM. User's Guide to Winsteps. Chicago: Mesa Press; 2010.
34. Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in Human Sciences (2nd ed). Mahwah, NJ: Lawrence Erlbaum; 2007:179-179.
35. Linacre JM. Structure in Rasch residuals: why principal components analysis (PCA). Rasch Meas Trans 1998;12(2):636 [FREE Full text]
36. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3(2):205-231. [Medline: 12011501]
37. Smith RM, Suh KK. Rasch fit statistics as a test of the invariance of item parameter estimates. J Appl Meas 2003;4(2):153-163. [Medline: 12748407]
38. Linacre JM. RUMM2020 Item-Trait Chi-Square and Winsteps DIF Size. Rasch Mea Trans 2007;21(1):1096 [FREE Full text]
39. Linacre JM. WINSTEPS. URL: http://www.winsteps.com/index.htm [accessed 2013-06-27] [WebCite Cache ID 6HgKbgGWF]
40. Patil VH, Surendra NS, Sanjay M, Donavan T. Parallel Analysis Engine to Aid Determining Number of Factors to Retain. 2007. URL: http://ires.ku.edu/~smishra/parallelengine.htm [accessed 2013-12-17] [WebCite Cache ID 6LuTpjiHo]
41. Patil VH, Singh SN, Mishra S, Todd Donavan D. Efficient theory development and factor retention criteria: Abandon the ‘eigenvalue greater than one’ criterion. Journal of Business Research 2008 Feb;61(2):162-170. [doi: 10.1016/j.jbusres.2007.05.008]
42. Linn RL. A Monte Carlo approach to the number of factors problem. Psychometrika 1968 Mar;33(1):37-71. [doi: 10.1007/BF02289675]
43. Silverstein AB. Comparison of two criteria for determining the number of factors. Psychological Reports 1977 Oct;41(2):387-390. [doi: 10.2466/pr0.1977.41.2.387]
44. Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. Psychological Bulletin 1986;99(3):432-442. [doi: 10.1037//0033-2909.99.3.432]
45. Eack SM, Singer JB, Greeno CG. Screening for anxiety and depression in community mental health: the beck anxiety and depression inventories. Community Ment Health J 2008 Dec;44(6):465-474. [doi: 10.1007/s10597-008-9150-y] [Medline: 18516678]
46. Ramirez Basco M, Bostic JQ, Davies D, Rush AJ, Witte B, Hendrickse W, et al. Methods to improve diagnostic accuracy in a community mental health setting. Am J Psychiatry 2000 Oct;157(10):1599-1605. [Medline: 11007713]
47. Shear MK, Greeno C, Kang J, Ludewig D, Frank E, Swartz HA, et al. Diagnosis of nonpsychotic patients in community clinics. Am J Psychiatry 2000 Apr;157(4):581-587. [Medline: 10739417]
48. Tsuno K, Kawakami N, Inoue A, Abe K. Measuring workplace bullying: reliability and validity of the Japanese version of the negative acts questionnaire. J Occup Health 2010;52(4):216-226 [FREE Full text] [Medline: 20571253]
49. Hsueh IP. Rasch analysis of combining two indices to assess comprehensive ADL function in stroke patients. Stroke 2004 Feb 12;35(3):721-726. [doi: 10.1161/01.STR.0000117569.34232.76]
50. Giorgi G, Arenas A, Leon-Perez JM. An operative measure of workplace bullying: the negative acts questionnaire across Italian companies. Ind Health 2011;49(6):686-695 [FREE Full text] [Medline: 22020012]
51. Nielsen MB, Notelaers G, Einarsen S. Measuring exposure to workplace bullying. In: Einarsen S , Hoel H , Zapf D , Cooper CL , editors. Bullying and Harassment in the Workplace: Developments in Theory, Research, and Practice. Boca Raton: Boca Raton CRC Press; 2011:149-176.
52. Leymann H. Mobbing and psychological terror at workplaces. Violence Vict 1990;5(2):119-126. [Medline: 2278952]

Abbreviations

AUROC: area under the ROC
CAT: computerized adaptive testing
CTT: classical test theory
DC: dimension coefficient
IRT: item response theory
MNSQ: mean square residual
SEM: standard error of the mean
NAQ-R: Negative Acts Questionnaire-Revised
NAT: nonadaptive testing
ROC: receiver operating characteristic

VBA: Visual Basic for Applications

Edited by G Eysenbach; submitted 09.07.13; peer-reviewed by CL Hsieh, C Lalanne; comments to author 29.08.13; revised version received 17.09.13; accepted 28.11.13; published 17.02.14.

Please cite as:
Ma SC, Chien TW, Wang HH, Li YC, Yui MS
Applying Computerized Adaptive Testing to the Negative Acts Questionnaire-Revised: Rasch Analysis of Workplace Bullying
J Med Internet Res 2014;16(2):e50
URL: *http://www.jmir.org/2014/2/e50/*
doi:*10.2196/jmir.2819*
PMID:*24534113*

Original Paper

Supporting Patients Treated for Prostate Cancer: A Video Vignette Study With an Email-Based Educational Program in General Practice

Moyez Jiwa[1], MB, BCh, BAO, FRCP, FRACGP, MRCGP, MD; Georgia Halkett[1], BMedRad, FIR, PhD; Xingqiong Meng[1], BMed, MAppE, PhD; Vinita Pillai[1], BSc, MSc; Melissa Berg[1], BSc; Tim Shaw[2], BSc, PhD

[1]Curtin University, Perth, Australia
[2]University of Sydney, Sydney, Australia

Corresponding Author:
Moyez Jiwa, MB, BCh, BAO, FRCP, FRACGP, MRCGP, MD
Curtin University
GPO Box U1987
Perth,
Australia
Phone: 61 8 9266 1768
Fax: 61 8 9266 2608
Email: m.jiwa@curtin.edu.au

Abstract

Background: Men who have been treated for prostate cancer in Australia can consult their general practitioner (GP) for advice about symptoms or side effects at any time following treatment. However, there is no evidence that such men are consistently advised by GPs and patients experience substantial unmet need for reassurance and advice.

Objective: The intent of the study was to evaluate a brief, email-based educational program for GPs to manage standardized patients presenting with symptoms or side effects months or years after prostate cancer treatment.

Methods: GPs viewed six pairs of video vignettes of actor-patients depicting men who had been treated for prostate cancer. The actor-patients presented problems that were attributable to the treatment of cancer. In Phase 1, GPs indicated their diagnosis and stated if they would prescribe, refer, or order tests based on that diagnosis. These responses were compared to the management decisions for those vignettes as recommended by a team of experts in prostate cancer. After Phase 1, all the GPs were invited to participate in an email-based education program (Spaced Education) focused on prostate cancer. Participants received feedback and could compare their progress and their performance with other participants in the study. In Phase 2, all GPs, regardless of whether they had completed the program, were invited to view another set of six video vignettes with men presenting similar problems to Phase 1. They again offered a diagnosis and stated if they would prescribe, refer, or order tests based on that diagnosis.

Results: In total, 64 general practitioners participated in the project, 57 GPs participated in Phase 1, and 45 in Phase 2. The Phase 1 education program was completed by 38 of the 57 (59%) participants. There were no significant differences in demographics between those who completed the program and those who did not. Factors determining whether management of cases was consistent with expert opinion were number of sessions worked per week (OR 0.78, 95% CI 0.67-0.90), site of clinical practice (remote practice, OR 2.25, 95% CI 1.01-5.03), number of patients seen per week (150 patients or more per week, OR 10.66, 95% CI 3.40-33.48), and type of case viewed. Completion of the Spaced Education did impact whether patient management was consistent with expert opinion (not completed, OR 0.88, 95% CI 0.5-1.56).

Conclusions: The management of standardized patients by GPs was particularly unlikely to be consistent with expert opinion in the management of impotence and bony metastasis. There was no evidence from this standardized patient study that Spaced Education had an impact on the management of patients in this context. However, the program was not completed by all participants. Practitioners with a greater clinical load were more likely to manage cases as per expert opinion.

(J Med Internet Res 2014;16(2):e63) doi:10.2196/jmir.3003

KEYWORDS

medical education; prostate cancer; general practice; email; video

Introduction

Prostate cancer has been the most commonly diagnosed cancer in Australian men since 1989 [1]. One in nine men in Australia will develop prostate cancer in their lifetime [2]. Most men with prostate cancer survive more than 5 years and die of unrelated causes [3]. The treatment of prostate cancer may include surgery, radiotherapy, chemotherapy, hormone treatment, or watchful waiting. Treatment depends on prognosis, stage of disease, treatment options, and side effects as well as the patient and his partner's preferences [4].

In the months and years following treatment, men may experience a number of troublesome side effects, or in the case of advanced disease, symptoms and signs related to metastatic disease. These include impotence, urinary incontinence, proctitis, depression, fatigue, and malignant bone pain [5]. Post-treatment follow-up is provided in the tertiary settings in some instances; however, this follow-up may only be for a short period of time after which patients are encouraged to see their general practitioner (GP) about any ongoing problems. Previous studies have demonstrated that men consult a GP routinely in the months and years after treatment for prostate cancer [6]. Prostate cancer patients are more likely to contact their GP for urinary problems and erectile dysfunction (ED) than for other symptoms [6]. GP presentation for fatigue is also more common in prostate cancer patients [6]. However, there is no evidence that such patients are appropriately advised by general practitioners, and patients experience substantial unmet need for reassurance and advice [7]. In order to address the needs of patients treated for prostate cancer, the general practitioner needs to be knowledgeable about the recommended treatment for side effects of radiation therapy and the signs and symptoms that merit urgent referral for further specialist treatment. There is some evidence that general practitioners require further education on the specific needs of men living with prostate cancer and especially those who have received radiation therapy [8,9].

Methods

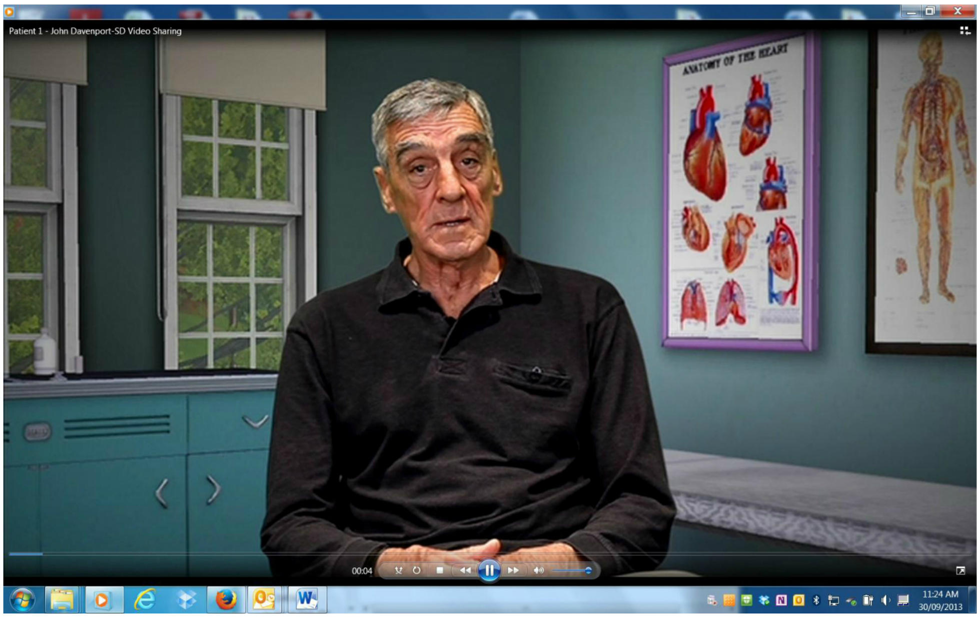
Participants

Following approval from the Curtin Human Research Ethics Committee (HR 08/2011), participants were recruited from a network of 150 GPs across Australia. GPs were emailed invitations and the initial form emails were supplemented with follow-up personal invitations to some of the 150 invitees who did not initially respond. Participants were remunerated with AUD $300 for their contribution.

Materials

Twelve video vignettes were developed, one pair for each potential side effect related to treatment for prostate cancer or the features of metastasis (see Multimedia Appendices 1 and 2 for exemplars). Each vignette depicted a patient with clear indications for specific management, including referral, prescription, reassurance, and/or investigation. The vignettes were developed by three GPs, a radiation oncologist, a medical oncologist, and a urologist. The expert panel also suggested the management of each case with details of prescription, referral for specialist treatment, and laboratory investigation (Multimedia Appendix 3). The vignettes were then prepared as a short video monologue by an actor-patient (Figure 1). The video included an off-camera commentary by an actor-doctor describing relevant signs to be found on clinical examination. Participation in the study was via the Internet. Participants were asked 4 questions after watching each video vignette: (1) "What is your diagnosis?", (2) "Would you prescribe something? If so, what?", (3) "Would you refer the patient? If so, to whom?", and (4) "Would you order tests? If so, which tests?"

Figure 1. Screenshot of video vignette.



Development of Spaced Education (Spaced Ed) Program

The program involves clinicians answering small numbers of case-based multiple choice questions that are emailed to them over a number of days. Participants receive one question at a time. Reponses can be submitted in one of two ways either by selecting from a choice of potential answers or by entering a 140 character free-text reply. The questions can be received by email or via a smartphone. The questions are repeated using an evidence-based adaptive algorithm that personalizes the spacing and content of a course to the demonstrated knowledge level of each learner. Learners receive succinct feedback after answering each case and can see how they are progressing through a program and compare their performance with peers. A program is completed when a participant answers all questions correctly twice consecutively. Feedback to the participants included references to the literature where the participants could read more about the subject if they wished. Completing a Spaced Education (Ed) program requires a few minutes every other day to answer the questions. Through the use of this technology, large numbers of geographically distributed practitioners can be reached relatively easily. Spaced Ed has previously been found in many studies to have an impact on knowledge and clinician behavior. A large randomized study investigating the impact of Spaced Ed on inappropriate PSA (prostate-specific antigen) screening by primary care clinicians in the Veterans Affairs Network in the United States found Spaced Ed significantly reduced inappropriate screening and this effect persisted for 2 years after the course [10]. A further randomized study in the United States found Spaced Ed impacted significantly on self-reported global clinical behaviors among primary care clinicians [11]. Spaced Ed is based on the spacing effect—the psychological finding that educational encounters that are spaced and repeated over time result in improved retention and more efficient learning compared to an educational event held at a single point in time [12,13].

The case studies and answers to the questions for the Spaced Ed program deployed in this study were developed by the same multidisciplinary team of clinicians involved in devising the video vignettes. The focus of the case studies matched the materials presented in the video vignettes as described above.

The project was completed in three stages: (1) Phase 1 - participants were invited to view the first set of 6 videos and to describe their management of the standardized patient depicted, (2) Spaced Education - all participants were invited to take part in the Spaced Ed program for 8 weeks, and (3) Phase 2 - all participating GPs were invited to view the second set of six videos and to describe their management of the standardized patient depicted.

Statistical Analysis

We estimated that approximately 50% of the participants would complete the Spaced Ed program in the specified 8 weeks. The proportion of those who managed cases as per the expert recommendations were expected to be greater in the Spaced Ed group (0.60 vs 0.30). For this reason, a sample of 42 participants per group was deemed sufficient in this exploratory study to estimate the effect size of the Spaced Ed program within 95% confidence intervals ranging from 0-34% [14].

Fisher’s exact tests were used to determine group differences in the proportion of cases diagnosed and managed correctly. Binary logistic regression was used to determine group differences in the correct management of cases by the participants. The full regression model included: the 2 participant groups; age; gender; country of graduation; years after graduation; years of GP experience; status as established GP or GP registrar (trainee); fellowship status with the Royal Australian College of General Practitioners (FRACGP); the remoteness of their primary practice; the number of GPs at their primary practice; status as a principal within their primary practice; patients seen per week; patient care hours per week; and whether they conduct non-English consultations. Regression models were constructed using both backwards elimination and forward selection. Variables with a *P* value less than .05 were retained in the final model and reported, with the exception of the variable of the intervention group, which remained in the model regardless of its significance level. Stata version 12.1 (StataCorp LP, College Station, TX, USA) was used to perform the analysis. Logistic regression models were adjusted for the lack of independence between individual participants by estimating the clustered standard errors to account for intra-group correlation (“vce” option in Stata).

Results

In total, 64 general practitioners consented to participate in the project, 57 GPs completed Phase 1, 45 completed Phase 2, and 38 of 57 (59%) participants completed the Spaced Education program. There were no significant differences in demographics between those who completed the Spaced Ed program and those who did not (Table 1). There were significant differences in the correct diagnosis of individual cases between the study’s two phases (Table 2). There were some statistically significant differences in the management of cases in Phase 1 compared to Phase 2 (Table 3). In Phase 2, there was no difference in the diagnosis of cases regardless of whether the participant had completed Spaced Ed or not (Table 4). Similarly, there was no difference in the management of cases, whether the participant had completed the Spaced Ed or not (Table 5).

Regression analysis was carried out to determine whether the GPs managed the case as recommended by experts with reference to three explanatory variables: (1) GP demographics, (2) Spaced Education, and (3) Cases. These variables explained 25% of the differences observed (R^2=.25) (Table 6). The number of sessions in general practice was strongly correlated with the number of patients seen per week (Pearson’s correlation coefficient of .74, $P<.001$). However, the number of sessions in practice were inversely correlated with correct case management. It is possible that some of the practitioners who were seeing a greater number of patients had received specialist training in prostate cancer; however, we could not test this hypothesis from these data. Male GPs did more clinical sessions (median 8, IQR 6) than females (median 6, IQR 6); however, gender did not have a significant influence in the regression analysis.

Table 1. Participant demographic information.

Characteristics	Total sample (n=64)	Spaced Ed, completed (n=38)	Spaced Ed, not completed (n=26)	*P* value
	mean (SD) or n (%)	mean (SD) or n (%)	mean (SD) or n (%)	
Participant demographics				
Age (years)	43.5 (11.3)	41.8 (10.7)	46.1 (11.9)	.14
Years after graduation	19.8 (11.5)	-	-	.11
Years of GP experience	15.0 (11.8)	13.2 (11.2)	17.5 (12.5)	.16
Number of other GPs at same clinic	7.4 (4.2)	6.6 (3.5)	8.5 (4.9)	.07
GP sessions worked/week	6.8 (2.8)	7.1 (2.7)	6.4 (3.0)	.36
Male	34 (53.1%)	18 (47.4%)	16 (61.5%)	.31
Graduated in Australia	48 (75.0%)	27 (71.1%)	21 (80.8%)	.56
Registrars	14 (21.9%)	7 (18.4%)	7 (26.9%)	.54
FRACGP[b]	34 (53.1%)	21 (55.3%)	13 (50.0%)	.80
Primary practice demographics				
Accredited	63 (98.4%)	38 (100.0%)	25 (96.2%)	.41
Location				.35[a]
ACT (Australian Capital Territory)	1 (1.6%)	1 (2.6%)	0 (0.0%)	
NSW (New South Wales)	10 (15.6%)	8 (21.1%)	2 (7.7%)	
QLD (Queensland)	5 (7.8%)	3 (7.9%)	2 (7.7%)	
SA (South Australia)	5 (7.8%)	3 (7.9%)	2 (7.7%)	
TAS (Tasmania)	1 (1.6%)	0 (0.0%)	1 (3.9%)	
VIC (Victoria)	14 (21.9%)	10 (26.3%)	4 (15.4%)	
WA (Western Australia)	26 (43.8%)	13 (34.2%)	15 (57.7%)	
Clinic remoteness				.93
Major city	44 (68.8%)	26 (68.4%)	18 (69.2%)	
Inner regional	9 (14.1%)	6 (15.8%)	3 (11.5%)	
Outer regional/remote	11 (17.2%)	6 (15.8%)	5 (19.2%)	
GP position				.13
Principal	15 (23.4%)	6 (15.8%)	9 (34.6%)	
Non-Principal	39 (60.9%)	27 (71.1%)	12 (46.1%)	
Others	10 (15.6%)	5 (13.2%)	5 (19.2%)	
Patient consultations				
Patient consultations per week				.80
<100	32 (50.0%)	18 (47.4%)	14 (53.9%)	
100-149	18 (28.2%)	12 (31.6%)	6 (23.1%)	
≥150	14 (21.9%)	8 (21.1%)	6 (23.1%)	
Patient consultation hours per week				1.00
<11	7 (10.9%)	4 (10.5%)	3 (11.5%)	
11-20	11 (17.2%)	7 (18.4%)	4 (15.4%)	
21-40	35 (54.7%)	21 (55.3%)	14 (53.9%)	
≥41	11 (17.2%)	6 (15.8%)	5 (19.2%)	
Non-English consultations				.53

Characteristics	Total sample (n=64) mean (SD) or n (%)	Spaced Ed, completed (n=38) mean (SD) or n (%)	Spaced Ed, not completed (n=26) mean (SD) or n (%)	*P* value
No	52 (81.3%)	32 (84.2%)	20 (76.9%)	
<25%	12 (18.8%)	6 (15.8%)	6 (23.1%)	

[a] *P* values were derived from Fisher's exact test.

[b] FRACGP: Fellowship Royal Australian College of General Practitioners

Table 2. Correct diagnosis of cases per phase of study.

Diagnosis		Phase 1, correct (n=57), n (%)	Phase 2, correct (n=45), n (%)	*P* value
Case 1				
	Radiation proctitis	42 (73.7%)	45 (100.0%)	<.001
Case 2				
	PSA bounce after radiation therapy	21 (36.8%)	26 (57.8%)	.04
Case 3				
	Spinal metastasis	50 (89.3%)	40 (88.9%)	.86
Case 4				
	Urethral stricture after radiotherapy	24 (43.6%)	38 (84.4%)	<.001
Case 5				
	Psychological cause	26 (47.3%)	23 (51.1%)	.58
	Adverse effect of medication	10 (18.2%)	19 (42.2%)	.006
Case 6				
	Biological depression	39 (70.9%)	10 (22.2%)	<.001
	Psychosocial factors	53 (96.4%)	45 (100.0%)	.07[a]

[a] *P* values were derived from Fisher's exact test.

Table 3. Correct management of cases by phase of study.

Management		Phase 1, correct (n=57), n (%)	Phase 2, correct (n=45), n (%)	*P* value
Case 1 (Proctitis)				
	Refer to specialist	15 (26.3%)	19 (42.2%)	.09
	Rule out bowel infection	17(29.8%)	21 (46.7%)	.08
	Refer for colonoscopy	30 (52.6%)	20 (44.4%)	.41
	Prescribe medication	1 (1.8%)	16 (35.6%)	<.001[a]
Case 2 (Anxiety)				
	No specific treatment	20 (35.1%)	17 (37.8%)	.77
	Reassure	4 (7.0%)	1 (2.2%)	.26[a]
Case 3 (Recurrence)				
	Refer to radiation oncologist	15 (26.8%)	17 (37.8%)	.22
	Seek specialist advice on investigations	36 (64.3%)	25 (55.6%)	.44
	Order plain x-rays	42 (75.0%)	32 (71.1%)	.77
Case 4 (Stricture)				
	Refer to urologist	19 (34.6%)	21 (46.7%)	.17
	Refer to physiotherapist	0 (0.0%)	1 (2.2%)	.25[a]
	Micturating cysto-urethrogram	27 (49.1%)	26 (57.8%)	.29
	Renal ultrasound scan	24 (43.6%)	15 (33.3%)	.36
Case 5 (Impotence)				
	Check cholesterol	13 (23.6%)	24 (53.3%)	.001
	Check blood glucose	14 (25.5%)	21 (46.7%)	.02
	Check hormone levels	12 (21.8%)	22 (48.9%)	.006
Case 6 (Depression)				
	Prescribe antidepressant	39 (73.6%)	40 (88.9%)	.01

[a] *P* values derived from Fisher's exact test.

Table 4. Correct diagnosis of cases as per completion of Spaced Ed (SE) in Phase 2 (n=45).

Phase 2		SE completed, n=37		SE not completed, n=8		*P* value[a]
		Diagnosed correctly	Diagnosed incorrectly	Diagnosed correctly	Diagnosed incorrectly	
Case 1						
	Radiation Proctitis	100.0	-	100.0	-	-
Case 2						
	PSA Bounce after radiation therapy	22	15	4	4	.70
Case 3						
	Spinal metastasis	33	4	7	1	1.00
Case 4						
	Urethral stricture after radiotherapy	32	5	6	2	.59
Case 5						
	Psychological cause	19	18	4	4	1.00
	Adverse effect of medication	15	22	4	4	.70
Case 6						
	Biological depression	10	27	0	8	.17
	Psychosocial factors	37	-	8	-	-

[a] All *P* values in this table were derived from Fisher's exact test.

Table 5. Correct management of cases as per completion of Spaced Ed (SE) in Phase 2 (n=45).

Phase 2	SE completed, n=37		SE not completed, n=8		
	Managed as recommended	Not managed as recommended	Managed as recommended	Not managed as recommended	*P* value[a]
Case 1 (Radiation)					
Refer to specialist	17	20	2	6	.43
Rule out bowel infection	18	19	3	5	.71
Refer for colonoscopy	16	21	4	4	1.0
Prescribe medication	-	-	-	-	-
Case 2 (PSA Bounce)					
No specific treatment	13	24	4	4	.45
Reassure	1	36	0	8	1.0
Case 3 (Recurrence)					
Refer to radiation oncologist	16	21	1	7	.13
Seek specialist advice on investigations	23	14	2	6	.11
Order plain x-rays	26	11	6	2	1.0
Case 4 (Stricture)					
Refer to urologist	18	19	3	5	.71
Refer to physiotherapist	1	36	0	8	1.0
Micturating cysto-urethrogram	23	14	3	5	.25
Ultrasound scan	14	23	1	7	.24
Case 5 (Impotence)					
Check cholesterol	21	16	3	5	.44
Check blood glucose	17	20	4	4	1.0
Check hormone levels	19	18	3	5	.7
Case 6 (Depression)					
Prescribe antidepressant	32	5	8	8	.27

[a] All *P* values derived from Fisher’s exact test other than cases where cells contained 5 or more participants where Pearson’s chi-square test was used.

Table 6. Regression analysis.

Variable		OR	95% CI	*P* value
Sessions worked per week		0.78	0.67-0.90	.001
GP registrar				
	Yes	1.00		
	No	4.66	2.23-9.71	<.001
Clinical remoteness				
	Major cities	1.00		
	Inner regional	0.38	0.21-0.70	.002
	Outer regional/remote	2.25	1.01-5.03	.048
Patients seen per week				
	<100	1.00		
	100-149	4.53	1.91-10.72	.001
	≥150	10.66	3.40-33.48	<.001
Spaced Education				
	Completed	1.00		
	Not completed	0.88	0.50-1.56	.66
Cases				
	1. Proctitis	1.00		
	2. PSA bounce	5.36	1.79-16.09	.003
	3. Bony metastasis	1.27	0.42-3.82	.67
	4. Urethral stricture	39.75	10.42-151.57	<.001
	5. Impotence	0.67	0.24-1.88	.44
	6. Depression	2.92	1.05-8.15	.04

Discussion

Principal Findings

Bowel, urinary or sexual dysfunction, depression, and anxiety are common presentations in primary care (1). In this study, patients with such problems were presented in the context of treatment for prostate cancer. Our data indicate that there were limited numbers of participants who correctly diagnosed the symptoms presented or suggested a management plan that was consistent with expert opinion. In Phase 1, the differences were marked for most cases (Tables 2 and 3). Such deviations from expert opinion have been reported previously [15,16]. From the regression analysis, we were able to conclude that compared to radiation proctitis, PSA bounce, urethral stricture, and depression were more likely to be managed as per the experts. However, erectile dysfunction was less likely to be managed as per expert opinion, especially in Phase 1. Erectile dysfunction is the most common side effect of treatment for early prostate cancer. It has far-reaching effects upon men’s lives [17]. Although some treatment effects such as radiation proctitis are relatively uncommon, ED is a common symptom that is likely to be presented to general practitioners in many clinical contexts [18].

Overall, the differences in management between the participants and the expert panel were less marked in Phase 2 of the study, and it is possible that in the intervening 8 weeks participants may have sought information on how to manage the adverse effects of prostate cancer treatment. The only exception was the diagnosis of biological depression, which seemed to deviate more from expert opinion in Phase 2. This was unexpected and it may have reflected a reticence to diagnose significant depression in that vignette and or it may be that the actor did not display the features of a significant depression in a way that persuaded more practitioners to come to that diagnosis.

The observation that rural practitioners were more attuned to expert management is consistent with the survey of Australian GPs, where rural GPs were more willing to be involved in providing supportive care to cancer patients than colleagues in metropolitan areas [19]. With respect to the main focus of our study, those who completed the Spaced Ed intervention were not more likely to concur with the expert panel. This is in contrast to other previous reported studies [10,11]. The reason for this is unclear, but may relate to the nature of the conditions being considered or the context in which this intervention was delivered. A primary care consultation is known to be complex, with a focus on the physical, social, and psychological components of any symptom or problem presented. This may not lend itself to an intervention that promotes the application of simple rules [20]. Regression analysis suggests that more influential variables impacting on the outcomes were some of the demographic characteristics of the participants; specifically, a greater clinical load. This was not unexpected for patients treated for prostate cancer because many of these problems are likely to present infrequently, and few doctors will have encountered them previously unless consulting a large number of men.

A number of approaches have been reported in the literature to promote consistent and reliable management of chronic conditions in primary care [21,22]. A few of these have focused specifically on the knowledge of general practitioners [23]. Data from our study suggests that focusing on knowledge alone may not be sufficient. A recent literature review reported that two other factors are also likely to be important in the context of a cancer diagnosis, namely, attitudes and beliefs [24,25]. These issues were not evaluated in this study. For example, we were unable to report the participants' attitude to the management of patients following treatment and whether they felt this role extended to investigating and treating conditions that may have resulted from specialist treatment [9]. The review of attitudes to this issue among Australian GPs suggests that there is a diversity of opinions on the matter [19]. Nor could we confirm that all participants had access to a radiation oncologist in their clinical practice and/or would have had the option to refer a patient with bony metastasis or radiation proctitis to such an expert. The available evidence suggests that this is not a safe assumption and that management plans would be impacted by the clinicians' experience in their local context [26]. Finally, we could not identify any practitioners who had any specialist training in prostate cancer. However, all participants were working as general practitioners when they participated in this study and it is reasonable to assume that there were a negligible number with specialist training in a specific cancer.

With respect to the format of education offered here, even though more than half the sample completed the Spaced Ed program, it was disappointing that more did not do so. We did not observe any significant differences in the demographic characteristics of those who completed the Spaced Ed program and those who did not. The program was based on email communication and relied on doctors checking emails on a regular basis. There is limited evidence from the Australian literature that general practitioners routinely deploy digital technology [27]. Second, there was no evidence that participants were satisfied that the answers provided to the emailed scenarios were consistent with what they considered best practice. Informal feedback suggested that the program was unpopular even among those who persevered with it. For example, some participants pointed out that they disagreed with the answers offered by the experts.

Conclusions

In this standardized patient study, Spaced Ed did not promote management plans that were consistent with expert opinion. While there was a marked improvement in the management of cases in Phase 2 of this study, this may be because participants were stimulated to seek information elsewhere on the management of such cases. Greater clinical load had a more significant and positive impact on the management of patients than the Spaced Ed. Further development of the Spaced Ed may be required before it can be tested again in the context of prostate cancer follow-up in general practice. Perhaps greater involvement of the target group of practitioners in setting the answers in Spaced Ed may be helpful. Alternatively, it is also possible that while Spaced Ed as a short, targeted educational program facilitated by information technology is attractive, it is unlikely to succeed.

Acknowledgments

This study was funded through the Priority-Driven Collaborative Cancer Research Scheme by the Australian Government with the support of Cancer Australia. We would like to thank the experts who assisted in developing the cases and advising on appropriate management.

Conflicts of Interest

None declared.

Authors' Contributions

MJ, GH, and TS designed the study; MJ, VP, and MB coordinated the study; MJ and XM analyzed the data; TS coordinated the educational components; and MJ, GH, XM, MB, TS, and VP drafted the paper.

Multimedia Appendix 1

Example of video.

[MOV File, 31MB - jmir_v16i2e63_app1.mov]

Multimedia Appendix 2

Typical cases.

[PDF File (Adobe PDF File), 266KB - jmir_v16i2e63_app2.pdf]

Multimedia Appendix 3

Specific recommendations for management of cases.

[PDF File (Adobe PDF File), 100KB - jmir_v16i2e63_app3.pdf]

References

1. Quinn M, Babb P. Patterns and trends in prostate cancer incidence, survival, prevalence and mortality. Part II: individual countries. BJU Int 2002 Jul;90(2):174-184. [Medline: 12081759]
2. Prostate Cancer Foundation of Australia. Prostate cancer statistics. 2012. URL: http://www.prostate.org.au/articleLive/pages/Prostate-Cancer-Statistics.html [accessed 2013-10-05] [WebCite Cache ID 6K7olf5Rw]
3. Parkin DM, Pisani P, Ferlay J. Global cancer statistics. CA: A Cancer Journal for Clinicians 1999 Jan 01;49(1):33-64. [doi: 10.3322/canjclin.49.1.33]
4. Davison BJ, Gleave ME, Goldenberg SL, Degner LF, Hoffart D, Berkowitz J. Assessing information and decision preferences of men with prostate cancer and their partners. Cancer Nurs 2002 Feb;25(1):42-49. [Medline: 11838719]
5. Schultheiss TE, Lee WR, Hunt MA, Hanlon AL, Peter RS, Hanks GE. Late GI and GU complications in the treatment of prostate cancer. Int J Radiat Oncol Biol Phys 1997 Jan 1;37(1):3-11. [Medline: 9054871]
6. Heins MJ, Korevaar JC, Rijken PM, Schellevis FG. For which health problems do cancer survivors visit their General Practitioner? Eur J Cancer 2013 Jan;49(1):211-218. [doi: 10.1016/j.ejca.2012.07.011] [Medline: 22897842]
7. O'Brien R, Rose PW, Campbell C, Weller D, Neal RD, Wilkinson C, Prostate Cancer Follow-up Group. Experiences of follow-up after treatment in patients with prostate cancer: a qualitative study. BJU Int 2010 Oct;106(7):998-1003. [doi: 10.1111/j.1464-410X.2010.09292.x] [Medline: 20230391]
8. Watson EK, O'Brien R, Campbell C, Weller D, Neal RD, Wilkinson C, Prostate Cancer Follow-Up Study Group. Views of health professionals on the role of primary care in the follow-up of men with prostate cancer. Fam Pract 2011 Dec;28(6):647-654 [FREE Full text] [doi: 10.1093/fampra/cmr034] [Medline: 21693683]
9. Halkett GK, Jiwa M, O'Shea C, Smith M, Leong E, Jackson M, et al. Management of cases that might benefit from radiotherapy: a standardised patient study in primary care. Eur J Cancer Care (Engl) 2012 Mar;21(2):259-265. [doi: 10.1111/j.1365-2354.2011.01314.x] [Medline: 22146103]
10. Kerfoot BP, Lawler EV, Sokolovskaya G, Gagnon D, Conlin PR. Durable improvements in prostate cancer screening from online spaced education a randomized controlled trial. Am J Prev Med 2010 Nov;39(5):472-478 [FREE Full text] [doi: 10.1016/j.amepre.2010.07.016] [Medline: 20965387]
11. Shaw T, Long A, Chopra S, Kerfoot BP. Impact on clinical behavior of face-to-face continuing medical education blended with online spaced education: a randomized controlled trial. J Contin Educ Health Prof 2011 Jun;31(2):103-108. [doi: 10.1002/chp.20113] [Medline: 21671276]
12. Glenberg AM, Lehmann TS. Spacing repetitions over 1 week. Mem Cognit 1980 Nov;8(6):528-538. [Medline: 7219173]
13. Toppino TC, Kasserman JE, Mracek WA. The effect of spacing repetitions on the recognition memory of young children and adults. J Exp Child Psychol 1991 Feb;51(1):123-138. [Medline: 2010724]
14. Hertzog MA. Considerations in determining sample size for pilot studies. Res Nurs Health 2008 Apr;31(2):180-191. [doi: 10.1002/nur.20247] [Medline: 18183564]
15. O’Brien R, Rose P, Campbell C, Weller D, Neal RD, Wilkinson C. “I wish I’d told them”: a qualitative study examining the unmet psychosexual needs of prostate cancer patients during follow-up after treatment. Patient Educ Couns 2011;84(2):200-207. [doi: 10.1016/j.pec.2010.07.006]
16. Halkett GK, Jiwa M, O'Shea C, Smith M, Leong E, Jackson M, et al. Management of cases that might benefit from radiotherapy: a standardised patient study in primary care. Eur J Cancer Care (Engl) 2012 Mar;21(2):259-265. [doi: 10.1111/j.1365-2354.2011.01314.x] [Medline: 22146103]
17. Bokhour BG, Clark JA, Inui TS, Silliman RA, Talcott JA. Sexuality after treatment for early prostate cancer. J Gen Intern Med 2001 Oct;16(10):649-655. [doi: 10.1111/j.1525-1497.2001.00832.x]
18. Prins J, Blanker MH, Bohnen AM, Thomas S, Bosch JL. Prevalence of erectile dysfunction: a systematic review of population-based studies. Int J Impot Res 2002 Dec;14(6):422-432 [FREE Full text] [doi: 10.1038/sj.ijir.3900905] [Medline: 12494273]
19. Johnson CE, Lizama N, Garg N, Ghosh M, Emery J, Saunders C. Australian general practitioners' preferences for managing the care of people diagnosed with cancer. Asia Pac J Clin Oncol 2012 Dec 26:1-9. [doi: 10.1111/ajco.12047] [Medline: 23279791]
20. Watkins C, Harvey I, Langley C, Gray S, Faulkner A. General practitioners' use of guidelines in the consultation and their attitudes to them. Br J Gen Pract 1999 Jan;49(438):11-15 [FREE Full text] [Medline: 10622009]
21. Epping-Jordan JE. Improving the quality of health care for chronic conditions. Quality and Safety in Health Care 2004 Aug 01;13(4):299-305. [doi: 10.1136/qshc.2004.010744]
22. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. N Engl J Med 2007 Jul 12;357(2):181-190. [doi: 10.1056/NEJMsr065990] [Medline: 17625132]

23. Moreau C. [Luxembourg: old problems, new solutions]. Rev Infirm 1990 Oct;40(15):27-30. [Medline: 2244127]
24. Jiwa MMA, Dadich A. Continuity of cancer care: Where do primary care practitioners fit in? Cancer Forum 2013;37(1):31-34 [FREE Full text]
25. Jiwa M, McManus A, Dadich A. The impact of knowledge, attitudes and beliefs on the engagement of primary and community-based healthcare professionals in cancer care: a literature review. Curr Med Res Opin 2013 Nov;29(11):1475-1482. [doi: 10.1185/03007995.2013.838154] [Medline: 23998506]
26. McConigley R, Holloway K, Smith J, Halkett G, Keyser J, Aoun S, et al. The diagnosis and treatment decisions of cancer patients in rural Western Australia. Cancer Nurs 2011;34(4):E1-E9. [doi: 10.1097/NCC.0b013e3181f795e7] [Medline: 20885301]
27. Sim MG, Khong E, Jiwa M. Does general practice Google? Aust Fam Physician 2008 Jun;37(6):471-474 [FREE Full text] [Medline: 18523704]

Abbreviations

ED: erectile dysfunction
GP: general practitioner
PSA: prostate-specific antigen test
Spaced Ed: spaced education

Edited by G Eysenbach; submitted 07.10.13; peer-reviewed by A Koskan, S Cutrona; comments to author 28.01.14; revised version received 30.01.14; accepted 31.01.14; published 26.02.14.

Please cite as:
Jiwa M, Halkett G, Meng X, Pillai V, Berg M, Shaw T
Supporting Patients Treated for Prostate Cancer: A Video Vignette Study With an Email-Based Educational Program in General Practice
J Med Internet Res 2014;16(2):e63
URL: http://www.jmir.org/2014/2/e63/
doi:10.2196/jmir.3003
PMID:24571952

Original Paper

Relative Validity of Micronutrient and Fiber Intake Assessed With Two New Interactive Meal- and Web-Based Food Frequency Questionnaires

Sara E Christensen[1], MSc; Elisabeth Möller[1], PhD; Stephanie E Bonn[1], MSc; Alexander Ploner[1], PhD; Olle Bälter[2], PhD; Lauren Lissner[3], PhD; Katarina Bälter[1], PhD

[1]Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Stockholm, Sweden
[2]KTH Royal Institute of Technology, School of Computer Science and Communication, Stockholm, Sweden
[3]University of Gothenburg, Department of Public Health and Community Medicine, Göteborg, Sweden

Corresponding Author:
Sara E Christensen, MSc
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics
Nobels väg 12a
Stockholm, SE-171 77
Sweden
Phone: 46 8 52482361
Fax: 46 8 314975
Email: sara.christensen@ki.se

Abstract

Background: The meal- and Web-based food frequency questionnaires, Meal-Q and MiniMeal-Q, were developed for cost-efficient assessment of dietary intake in epidemiological studies.

Objective: The objective of this study was to evaluate the relative validity of micronutrient and fiber intake assessed with Meal-Q and MiniMeal-Q. The reproducibility of Meal-Q was also evaluated.

Methods: A total of 163 volunteer men and women aged between 20 and 63 years were recruited from Stockholm County, Sweden. Assessment of micronutrient and fiber intake with the 174-item Meal-Q was compared to a Web-based 7-day weighed food record (WFR). Two administered Meal-Q questionnaires were compared for reproducibility. The 126-item MiniMeal-Q, developed after the validation study, was evaluated in a simulated validation by using truncated Meal-Q data.

Results: The study population consisted of approximately 80% women (129/163) with a mean age of 33 years (SD 12) who were highly educated (130/163, 80% with >12 years of education) on average. Cross-classification of quartiles with the WFR placed 69% to 90% in the same/adjacent quartile for Meal-Q and 67% to 89% for MiniMeal-Q. Bland-Altman plots with the WFR and the questionnaires showed large variances and a trend of increasing underestimation with increasing intakes. Deattenuated and energy-adjusted Spearman rank correlations between the questionnaires and the WFR were in the range ρ=.25-.69, excluding sodium that was not statistically significant. Cross-classifications of quartiles of the 2 Meal-Q administrations placed 86% to 97% in the same/adjacent quartile. Intraclass correlation coefficients for energy-adjusted intakes were in the range of .50-.76.

Conclusions: With the exception of sodium, this validation study demonstrates Meal-Q and MiniMeal-Q to be useful methods for ranking micronutrient and fiber intake in epidemiological studies with Web-based data collection.

(J Med Internet Res 2014;16(2):e59) doi:10.2196/jmir.2965

KEYWORDS

validity; reproducibility; FFQ; micronutrients; weighed food record; Internet; adult

Introduction

The increasing use of the Internet worldwide [1] has made Web-based food frequency questionnaire (FFQ) methodology an attractive alternative to traditional paper-based instruments in epidemiological research. Today, more than 90% of the Swedish adult population has Internet access [2], which is a convincing rationale for choosing the Web over the paper-and-pencil method. Compared to paper-based FFQs, expenses are dramatically lower for Web-based versions for both dissemination and data handling, making it a more cost-efficient method [3,4]. In addition, the costs for the required software infrastructure have decreased over recent years [5]. The dynamic nature of the Web enables an interactive design with follow-up questions and skip patterns, which adapts the questions to the respondent's answers thereby reducing the answering time. An interactive Web-questionnaire has previously shown high compliance in a Swedish population with widespread Internet access [6]. Taking advantage of the benefits of using the Web, we have developed 2 Web-based FFQs with an interactive design; Meal-Q and MiniMeal-Q. The questionnaires have a meal-based format to ease recall of food intake. This approach has shown promising results in previous studies when compared with traditional food group designs [7,8].

We have previously published results on the validity of energy and macronutrient intake assessed by Meal-Q and MiniMeal-Q with doubly labeled water and a weighed food record (WFR) as reference methods [9]. The present paper evaluates the validity of micronutrient and fiber intake assessed by Meal-Q and MiniMeal-Q by using the WFR as the reference method. We also present an evaluation of the reproducibility of Meal-Q.

Methods

Background

Meal-Q was developed with guidance from a population-based cross-sectional study on food products consumed for breakfast, lunch, dinner, and snack meals as reported by 700 randomly selected Swedish participants through either face-to-face interviews or 24-hour telephone recalls (E Möller and S Christensen, written communication, August 2008). In the spring of 2009, Meal-Q was evaluated in the VALidation of Methods Assessing diet and physical activity (VALMA) validation study. The reference method was a 7-day WFR on the Web. After the VALMA study was completed, we developed the shorter version, MiniMeal-Q, by omitting food items with low consumption frequency and low contribution to total energy and nutrient intake. However, food items that were important sources of specific nutrient intakes were kept (eg, black pudding, which contributes to iron intake). Moreover, varieties of similar food items were also kept to enable analyses of dietary patterns (eg, different types of bread and cereals with varying fiber and sugar content). By using truncated data from Meal-Q, we simulated a validity evaluation of MiniMeal-Q. Acknowledging that MiniMeal-Q originated from Meal-Q data, their validity comparison should be interpreted with caution.

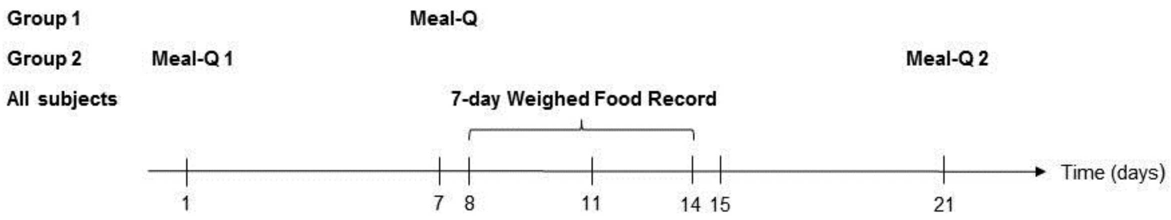
Recruitment

A total of 180 healthy volunteer men and women aged between 20 and 63 years were recruited through public announcement in Stockholm County, Sweden, to participate in the VALMA study. Announcements were made in the city, the suburbs, and at 2 universities, including among students in nutrition. Prerequisites for eligibility were access to the Internet and an email address, as well as not being on a weight-loss diet, nor being pregnant or having given birth during the past 10 months. All participants were informed about the study at an introductory meeting and gave their written informed consent. The Research Ethics committee at Karolinska Institutet approved the study.

Study Design

A study scheme of the 3-week validation study is shown in Figure 1. Participants were divided into 2 groups balanced for gender and age: group 1 (n=87) and group 2 (n=93). Group 1 filled out Meal-Q once, whereas group 2 also filled out a second Meal-Q after 3 weeks. Validity analysis with the WFR was made by using data from each participant's first-administered Meal-Q. For the simulated validity analysis of MiniMeal-Q, data from the first Meal-Q from both groups was truncated and compared to the WFR. For reproducibility analysis, the first and second Meal-Q from group 2 were compared. The first-administered Meal-Q additionally included questions on education, occupation, and tobacco use (current smoking and Swedish snuff use). Each participant self-reported their height and weight, which were used to calculate body mass index (BMI, kg/m^2).

Figure 1. The 3-week study scheme of the VALidation of Methods Assessing diet and physical activity (VALMA) study. Data from the first administered Meal-Q from both groups was compared to the WFR for validity analyses. The same data from Meal-Q was truncated for simulated validity analysis of MiniMeal-Q. Meal-Q was distributed twice in group 2 for reproducibility analysis.

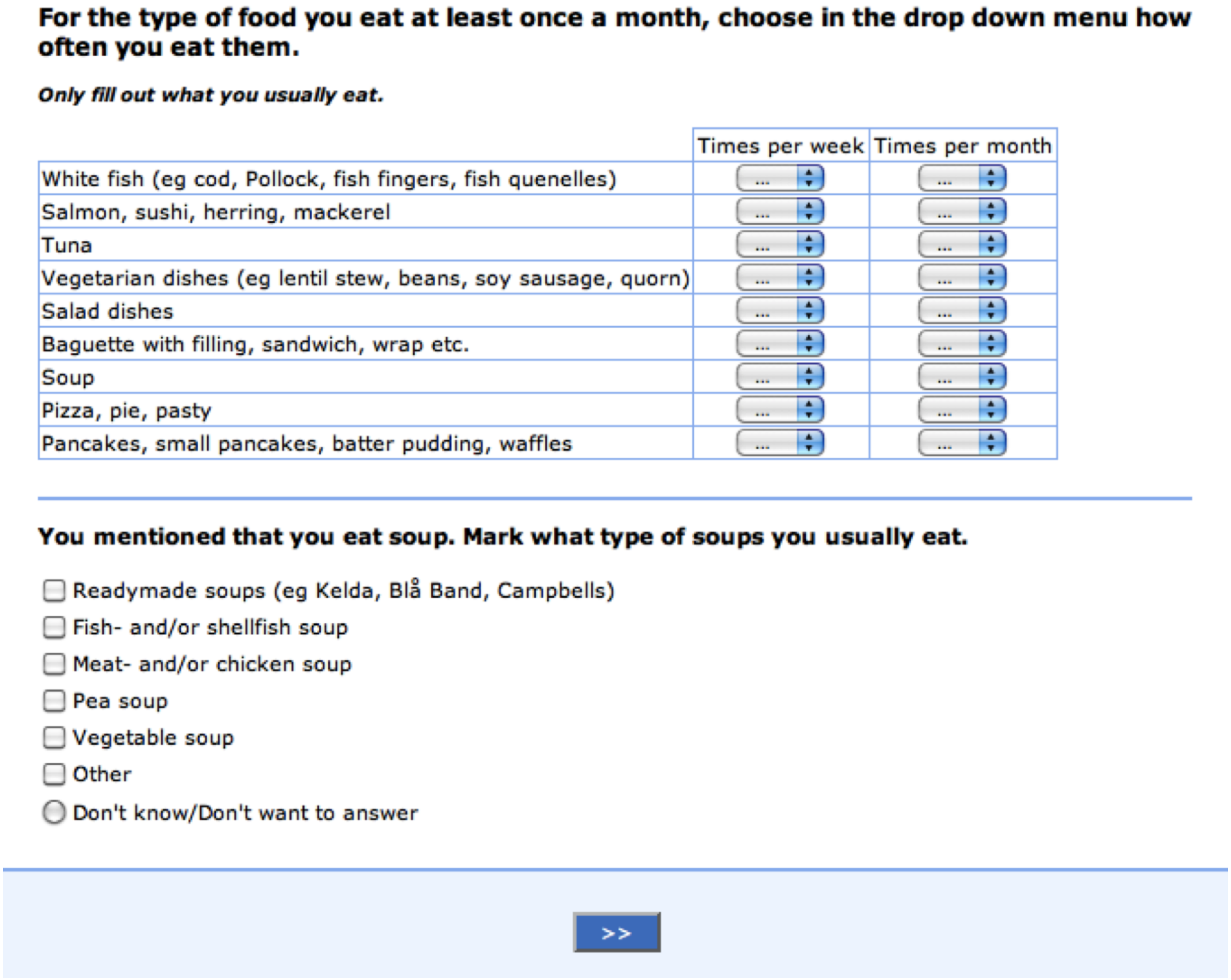


Dietary Assessment

Meal-Q

Meal-Q is interactive and includes 102 to 174 food items depending on the number of follow-up questions (see Figure 2 for an example of a questionnaire module). The mean answering time was 17 minutes (SD 11) in the current study population [9]. The interactivity implies follow-up questions for high consumers of certain food items and dishes. Meal-Q assesses intake of food items, dishes, and beverages, which enables the calculation of energy and nutrient intake (including alcohol). It also asks about meal patterns; eating behavior, such as restaurant visits; intake of fast food, light products, probiotics, the use of cooking fat and salt, as well as the use of dietary supplements. Respondents choose among predefined food items and intake frequencies ranging from 1-3 times a month to 5+ times a day. Five photos of portion sizes are included for each of the following food groups: (1) rice/potatoes/pasta, (2) meat/chicken/fish/vegetarian alternatives, and (3) vegetables (raw or cooked). The photos are used to calculate portion sizes for cooked dishes and vegetables, whereas standard portion sizes are used for other food items. The standard portion sizes are derived from the National Food Agency, the Swedish Consumer Agency, measured portion sizes developed by the research group, and standard portion sizes used in other FFQs at Karolinska Institutet. For this validation study, Meal-Q asked about dietary intake during the past few months.

Figure 2. Screenshot of a Meal-Q module: lunch and dinner dishes and a follow-up question on soup. Translated from the Swedish questionnaire version in the VALidation of Methods Assessing diet and physical activity (VALMA) study.



For the type of food you eat at least once a month, choose in the drop down menu how often you eat them.

Only fill out what you usually eat.

	Times per week	Times per month
White fish (eg cod, Pollock, fish fingers, fish quenelles)	...	...
Salmon, sushi, herring, mackerel	...	...
Tuna	...	...
Vegetarian dishes (eg lentil stew, beans, soy sausage, quorn)	...	...
Salad dishes	...	...
Baguette with filling, sandwich, wrap etc.	...	...
Soup	...	...
Pizza, pie, pasty	...	...
Pancakes, small pancakes, batter pudding, waffles	...	...

You mentioned that you eat soup. Mark what type of soups you usually eat.

☐ Readymade soups (eg Kelda, Blå Band, Campbells)
☐ Fish- and/or shellfish soup
☐ Meat- and/or chicken soup
☐ Pea soup
☐ Vegetable soup
☐ Other
○ Don't know/Don't want to answer

MiniMeal-Q

MiniMeal-Q contains 75 to 126 food items and is identical to Meal-Q in its design, including the interactive feature with adapted follow-up questions. The mean answering time for MiniMeal-Q was 7 minutes (SD 4) in a subsample of the current study population [9].

Weighed Food Record on the Web

At the introductory meeting, participants were given oral instructions and a handbook about how to fill out the 7-day WFR using a Web-based program, which covered more than 2000 food items. Each participant was given a kitchen scale and was asked to weigh and report all consumed food products and beverages at the highest level of detail possible. For example, a dish was encouraged to be reported by its individually weighed food items. As help for the recording throughout the day, all participants were provided with paper diaries. All records in the Web-based program were checked for completeness and reasonableness by the data collectors.

Nutrient Database

Daily intake of micronutrients and fiber was retrieved by linking intake of food items and dishes assessed with Meal-Q, MiniMeal-Q, and the WFR to the national database on nutrient content published by the Swedish National Food Agency [10]. The questionnaire's nutrient conversion was made by computer programs (MealCalc and MiniMealCalc) developed and validated by the research group specifically for this study. The nutrient conversion of the food items and dishes assessed with the WFR was built into the Web-based WFR program. The nutrient conversions did not include dietary supplements.

Assessment of Physical Activity Level for Identification of Energy Underreporters

A 7-day pedometer-based physical activity record provided in conjunction with the WFR program was filled out by all participants. The information was used to calculate the physical activity level (PAL) for each participant. Individual PAL values were also obtained from measurements of energy expenditure by the doubly labeled water (DLW) method [11] for 39 participants in group 2. A detailed description of the use of the DLW method in the VALMA study has been published previously [9,12]. The PAL values derived from the pedometers and from the DLW measurements were used for identification of potential energy underreporters in the WFR to exclude them from the comparison with Meal-Q and MiniMeal-Q.

Statistical Analysis

Descriptive characteristics of the study participants are presented as mean (SD) and as counts (%). A 2-sample *t* test was used to assess differences in BMI and age between study groups, between men and women, and between included and excluded participants. Fisher’s exact test was used to assess differences in education, nutrition background (studying or working in the nutrition field), and tobacco use. All tests were 2-sided with significance level alpha=.05.

Median (interquartile range, IQR) crude micronutrient and fiber intake was calculated and compared between Meal-Q, MiniMeal-Q, and the WFR, and differences between the methods were determined by using Wilcoxon signed rank tests. Linear regression was used to calculate the between-person variance captured in the truncated MiniMeal-Q as compared to Meal-Q. Identification of energy underreporters was made by using the Goldberg cut-off method [13]. The cut-off was calculated by using the energy intake from the WFR together with the obtained PAL values from the physical activity record and the DLW data.

For validity and reproducibility analyses, micronutrient and fiber intakes were adjusted for total energy intake by using the residual method [14]. To test the ranking agreement and magnitude of misclassification of the questionnaires in comparison to the WFR, we used quartile cross-classifications, calculating proportions of participants classified into the same, adjacent, and extreme quartiles of energy-adjusted intakes. Bland-Altman plots were presented for Meal-Q, MiniMeal-Q, and the WFR to evaluate absolute agreement and differences in bias within the intake range [15]. The differences between the questionnaires and the WFR were plotted against the mean of the 2 methods and the degree of variation was represented by the limits of agreement (ie, ±2 SDs of the mean difference). Most variables were not normally distributed after energy adjustments; therefore, Spearman rank correlation coefficients were used to compare the questionnaires to the WFR. Deattenuated correlations corrected for within-person variation in the WFR were calculated by using the formula of Beaton et al [16] and Liu et al [17] and confidence intervals were produced by using the method of Willett and Rosner [18].

Reproducibility of Meal-Q was evaluated by comparing crude median (IQR) micronutrient and fiber intake between the first and second Meal-Q and by cross-classification of energy-adjusted [14] quartiles. Intraclass correlation coefficients (ICCs) [19] were also computed by using 1-way ANOVA with random effects.

Statistical analyses were performed using STATA statistical software version 11.2 (StataCorp LP, College Station, TX, USA).

Results

Exclusions

One participant was excluded because of dropout (group 1) and 2 because of illness (group 2). With the Goldberg cut-off method, 14 participants (4 in group 1, 10 in group 2) were identified as energy underreporters in the WFR and were excluded. Hence, 163 participants remained for validity analysis (group 1: n=82; group 2: n=81). We found no significant differences between included and excluded participants in terms of age, BMI, education, nutrition background, or tobacco use (*P*=.16-.99). In the WFR assessments, 1 participant had implausibly high intakes of beta-carotene (>30,000 µg/day) and 3 other participants had implausibly high intakes of sodium (>9000 mg/day). Therefore, they were excluded in each respective nutrient analysis. For reproducibility analysis of Meal-Q, no exclusion of energy underreporters were made; however, 4 participants had missing values in the second administered Meal-Q, leaving 87 participants in the analysis.

Descriptive Statistics

General characteristics of the study participants included in the validity analysis are shown in Table 1. Most participants were highly educated (130/163, 80%) or students (95/163, 58%), one-third (54/163, 33%) were working full time, nearly one-third (49/163, 30%) had a nutrition background, and few participants (21/163, 13%) used tobacco. There were no statistically significant differences between study groups or sexes regarding age, BMI, education, nutrition background, smoking, or multivitamin/mineral supplement use (*P*=.05-.99). However, more men than women used Swedish snuff (*P*=.001). The between-person variance in micronutrient and fiber intake captured by MiniMeal-Q as compared to Meal-Q was 70% to 100%.

Table 1. Characteristics[a] of participants included in the validity analysis (N=163).

Characteristics	Group 1 (n=82)	Group 2 (n=81)	Men (n=34)	Women (n=129)	All (N=163)
Sex, n (%)					
Male	16 (20)	18 (22)			
Female	66 (80)	63 (78)			
Age (years), mean (SD)	34 (12)	32 (11)	33 (10)	33 (12)	33 (12)
BMI (kg/m^2), mean (SD)	23 (4)	23 (4)	24 (2)	23 (4)	23 (4)
Education >12 years, n (%)	64 (78)	66 (81)	27 (79)	103 (80)	130 (79.8)
Working full time, n (%)	33 (40)	21 (26)	12 (35)	42 (33)	54 (33.1)
Student, n (%)	41 (50)	54 (67)	18 (53)	77 (60)	95 (58.3)
Nutrition background,[b] n (%)	21 (26)	28 (35)	6 (18)	43 (33)	49 (30.1)
Tobacco use,[c] n (%)	11 (13)	10 (12)	12 (35)	9 (7)	21 (12.9)
Multivitamin/mineral supplement used, n (%)	18 (22)	14 (17)	8 (24)	24 (19)	32 (19.6)

[a] There was no statistically significant difference in characteristics between groups or sexes (*P*=.05-.99), except for Swedish snuff between sexes (1.8% women and 4.2% men, *P*=.001) using 2-sample *t* test and Fisher's exact test.

[b] Studying or working in the nutrition field.

[c] Tobacco use=current smoking and/or Swedish snuff use. Values are missing for 3 women in group 2.

[d] Daily or weekly supplement use assessed with Meal-Q.

Validity

The median (IQR) intake of most nutrients was higher when assessed with the WFR than with Meal-Q and MiniMeal-Q (Table 2). Exceptions were beta-carotene intake, which was higher when assessed with Meal-Q, whereas the intake was comparable between the WFR and MiniMeal-Q. There were no differences between the WFR and Meal-Q for thiamine, folate, magnesium, and fiber intake. Nor were there any differences between Meal-Q and MiniMeal-Q regarding any of the nutrients.

Quartile cross-classifications of micronutrient and fiber intake with the WFR and the questionnaires (Table 3) placed 69% to 90% of the participants into the same or adjacent quartile for Meal-Q, with the highest-ranking agreement for fiber and the lowest for sodium. For MiniMeal-Q, the ranking agreement ranged from 67% to 89%, also with fiber having the highest and sodium the lowest agreement. Proportions of participants in the extreme quartile ranged from 1% to 10% for Meal-Q and from 3% to 11% for MiniMeal-Q; the lowest proportions were found for vitamin C, magnesium, and fiber, and the highest for sodium.

Table 2. Median (IQR) daily crude micronutrient and fiber intake[a] assessed with the weighed food record (WFR), Meal-Q, and MiniMeal-Q (N=163).

Nutrients	WFR, median (IQR)	Meal-Q, median (IQR)	MiniMeal-Q, median (IQR)
Beta-carotene (µg)	2632 (2539)[b]	3372 (2905)	3254 (3079)
Thiamine (mg)	1.5 (0.5)	1.4 (0.8)	1.3 (0.8)
Riboflavin (mg)	1.9 (0.7)	1.7 (0.9)	1.5 (0.8)
Niacin (mg)	36 (14)	15 (9)	14 (8)
Vitamin B_6 (mg)	2.3 (1.0)	1.8 (1.0)	1.7 (0.9)
Folate (µg)	334 (167)	315 (210)	289 (193)
Vitamin B_{12} (µg)	5.7 (3.7)	3.8 (2.4)	3.5 (2.3)
Vitamin C (mg)	121 (92)	101 (82)	94 (74)
Vitamin D (µg)	5.6 (4.0)	4.7 (3.3)	4.4 (3.1)
Vitamin E (mg)	11 (5)	10 (5)	9 (5)
Calcium (mg)	1052 (381)	897 (583)	828 (512)
Iron (mg)	13 (6)	13 (7)	11 (6)
Magnesium (mg)	413 (177)	397 (242)	358 (207)
Phosphorus (mg)	1570 (514)	1433 (731)	1305 (677)
Potassium (mg)	3437 (1332)	3130 (1600)	2837 (1477)
Selenium (µg)	45 (21)	44 (24)	36 (22)
Sodium (mg)	3194 (1212)[c]	2448 (1118)	2158 (1015)
Zinc (mg)	12 (4)	11 (5)	10 (5)
Fiber (g)	25 (15)	26 (20)	23 (18)

[a] Most nutrient intakes assessed with the WFR were higher than intakes assessed with Meal-Q and MiniMeal-Q (*P*<.001-.03). Exceptions were beta-carotene intake, which was assessed higher with Meal-Q (*P*=.03), but was similar comparing the WFR and MiniMeal-Q (*P*=.19). Thiamine, folate, magnesium, and fiber intake were similar between the WFR and Meal-Q (*P*=.16-.92). There was no difference in intakes between Meal-Q and MiniMeal-Q (*P*<.001). (Wilcoxon signed rank test).

[b] n=162 because of exclusion of 1 participant with implausibly high intake.

[c] n=160 because of exclusion of 3 participants with implausibly high intakes.

Table 3. Quartile cross-classifications of mean daily energy-adjusted micronutrient and fiber intake assessed with Meal-Q, MiniMeal-Q, and the weighed food record (WFR) (N=163).

Nutrients	Same quartile, %		Adjacent quartile, %		Same/adjacent quartile, %		Extreme quartile, %	
	Meal-Q	MiniMeal-Q	Meal-Q	MiniMeal-Q	Meal-Q	MiniMeal-Q	Meal-Q	MiniMeal-Q
Beta-carotene (µg)[a]	41	41	40	42	81	83	4	6
Thiamine (mg)	27	31	44	43	71	74	7	5
Riboflavin (mg)	37	36	37	40	74	76	4	4
Niacin (mg)	36	34	45	45	81	79	4	5
Vitamin B_6 (mg)	34	31	41	45	75	76	6	6
Folate (µg)	42	40	38	42	80	82	4	4
Vitamin B_{12} (µg)	44	39	34	33	78	72	5	4
Vitamin C (mg)	39	38	46	45	85	83	2	3
Vitamin D (µg)	36	35	40	40	76	75	7	6
Vitamin E (mg)	40	42	34	33	74	75	4	4
Calcium (mg)	36	35	38	36	74	71	7	9
Iron (mg)	38	37	41	40	79	77	4	5
Magnesium (mg)	42	39	40	44	82	83	3	3
Phosphorus (mg)	33	34	42	43	75	77	7	7
Potassium (mg)	36	37	42	42	79	79	5	6
Selenium (µg)	41	38	37	42	78	80	4	6
Sodium (mg)[b]	33	35	36	32	69	67	10	11
Zinc (mg)	34	33	43	43	77	77	7	7
Fiber (g)	53	55	37	34	90	89	1	3

[a] n=162 because of exclusion of 1 participant with implausibly high intake.

[b] n=160 because of exclusion of 3 participants with implausibly high intakes.

The Bland-Altman plots with the WFR were similar for Meal-Q and MiniMeal-Q as seen in Table 4 and Figure 3 (showing an example of 8 micronutrients) and Multimedia Appendices 1-4. Niacin was largely underestimated by approximately 20 mg for both questionnaires. Most nutrients showed increasing underestimation with increasing intakes, and some also had a trend of increasing variance at higher intakes. In contrast, fiber had a larger variance at lower compared to higher intakes. Most of the nutrients had a varying bias over the intake range (ie, both underestimation and overestimation of intake with a magnitude approximately the same size of the mean intake). However, zinc, magnesium, potassium, and phosphorus showed a less varying bias.

Table 5 shows the Spearman correlation coefficients between Meal-Q, MiniMeal-Q, and the WFR. Correlations for Meal-Q for crude intakes were in the range ρ=.16-.66. Excluding the statistically nonsignificant correlation for sodium, the energy-adjusted correlations for Meal-Q ranged from ρ=.28-.67 and the deattenuated correlations ranged from ρ=.31-.69. The correlations were very similar for MiniMeal-Q, except for thiamine, which showed a stronger correlation with MiniMeal-Q than with Meal-Q.

Table 4. Overview of results from Bland-Altman plots[a] of Meal-Q and MiniMeal-Q in comparison with the weighed food record (WFR) (n=163).

Nutrients	Meal-Q		MiniMeal-Q		Meal-Q and MiniMeal-Q trends[b]
	Mean difference	±2 SD	Mean difference	±2 SD	
Beta-carotene (µg)[c]	427	–4100, 4985	285	–4300, 4873	Increasing variance with increasing intakes
Thiamine (mg)	0.01	–1.6, 16	–0.1	–1.8, 1.5	Increasing variance with increasing intakes
Riboflavin (mg)	–0.2	–1.3, 0.8	–0.3	–1.3, 0.7	Increasing underestimation with increasing intakes
Niacin (mg)	–21	–36, –5	–22	–37, –7	Increasing underestimation with increasing intakes
Vitamin B_6 (mg)	–0.4	–1.7, 0.9	–0.6	–1.9, 0.7	Increasing underestimation and variance with increasing intakes
Folate (µg)	–15	–245, 215	–50	–280, 180	Increasing underestimation with increasing intakes
Vitamin B_{12} (µg)	–2.0	–7.6, 3.6	–2.5	–8.0, 3.0	Increasing underestimation with increasing intakes
Vitamin C (mg)	–21	–142, 99	–29	–151, 93	Increasing underestimation and variance with increasing intakes
Vitamin D (µg)	–1.3	–9.4, 6.8	–1.6	–9.7, 6.6	Increasing underestimation with increasing intakes
Vitamin E (mg)	–1.4	–9.8, 6.9	–2.0	–10.0, 6.4	Increasing underestimation and variance with increasing intakes
Calcium (mg)	–113	–803, 576	–183	–892, 526	Increasing variance with increasing intakes
Iron (mg)	–1.0	–10.0, 8.0	–2.5	–12.0, 6.7	Increasing underestimation and variance with increasing intakes
Magnesium (mg)	–7.5	–206.0, 191.0	–41	–244, 162	Increasing underestimation with increasing intakes
Phosphorus (mg)	–164	–779, 450	–291	–904, 322	Increasing underestimation with increasing intakes
Potassium (mg)	–315	–1800, 1180	–640	–2200, 878	Increasing underestimation with increasing intakes
Selenium (µg)	–4	–40, 32	–10	–46, 25	Increasing underestimation with increasing intakes
Sodium (mg)[d]	–753	–2700, 1238	–1000	–3000, 922	Increasing underestimation and variance with increasing intakes
Zinc (mg)	–1.0	–6.5, 4.4	–1.9	–7.3, 3.4	Increasing variance with increasing intakes
Fiber (g)	2.0	–16.0, 20.0	–1.5	–20.0, 17.0	Larger variance at lower intakes than at higher intakes

[a] The Bland-Altman plots are shown in Figure 3 and Multimedia Appendices 1-4.

[b] Trends are similar for Meal-Q and MiniMeal-Q.

[c] n=162 because of exclusion of 1 participant with implausibly high intake.

[d] n=160 because of exclusion of 3 participants with implausibly high intakes.

Table 5. Spearman rank correlation coefficients (ρ) between Meal-Q, MiniMeal-Q, and the weighed food record (WFR) (N=163).

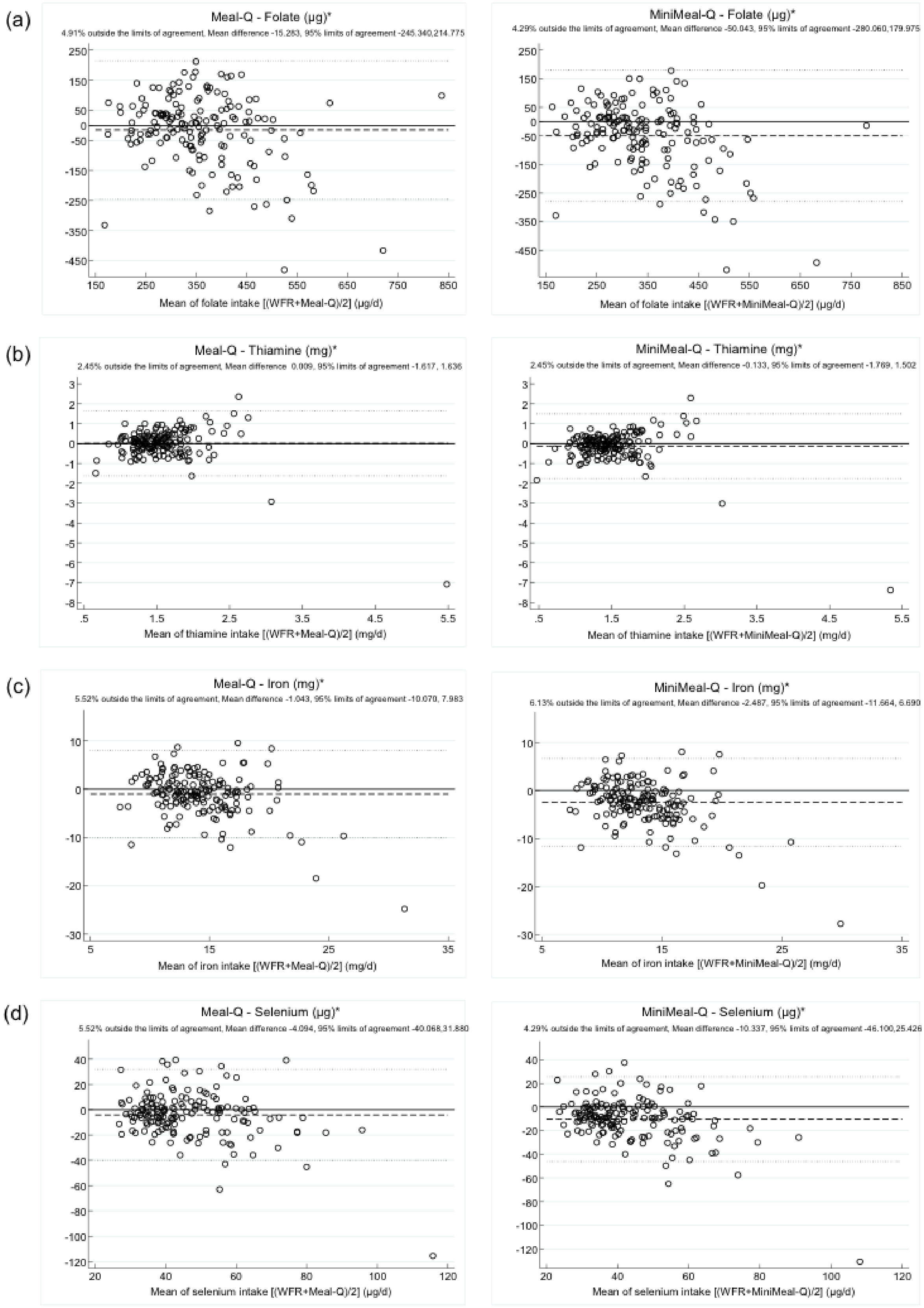
Nutrients	Crude ρ		Energy-adjusted ρ		Deattenuated ρ (95% CI)	
	Meal-Q	MiniMeal-Q	Meal-Q	MiniMeal-Q	Meal-Q	MiniMeal-Q
Beta-carotene[b]	.51	.51	.46	.46	.51 (.36, .64)	.51 (.36, .64)
Thiamine	.33	.37	.28	.35	.35 (.16, .52)	.43 (.24, .59)
Riboflavin	.16	.15	.39	.38	.42 (.27, .55)	.41 (.26, .54)
Niacin	.39	.37	.43	.42	.47 (.32, .59)	.46 (.31, .59)
Vitamin B_6	.39	.40	.32	.32	.35 (.19, .49)	.35 (.19, .49)
Folate	.50	.50	.50	.50	.53 (.39, .64)	.53 (.39, .64)
Vitamin B_{12}	.39	.28	.46	.37	.51 (.36, .63)	.41 (.26, .55)
Vitamin C	.53	.52	.53	0·50	.57 (.36, .64)	.54 (.41, .65)
Vitamin D	.34	.32	.31	.30	.36 (.19, .50)	.34 (.18, .49)
Vitamin E	.30	.30	.42	.42	.48 (.32, .61)	.48 (.33, .61)
Calcium	.23	.22	.29	.24	.31 (.16, .45)	.25 (.09, .40)
Iron	.44	.43	.42	.38	.46 (.31, .59)	.42 (.27, .55)
Magnesium	.52	.52	.54	.52	.56 (.44, .66)	.54 (.41, .64)
Phosphorus	.36	.37	.36	.36	.39 (.24, .52)	.39 (.24, .52)
Potassium	.42	.42	.41	.38	.43 (.29, .56)	.40 (.25, .52)
Selenium	.32	.30	.42	.41	.45 (.30, .57)	.44 (.30, .57)
Sodium[c]	.32	.32	.15	.12	.16 (-.01, .32)	.14 (-.04, .30)
Zinc	.33	.34	.31	.31	.34 (.18, .49)	.35 (.19, .49)
Fiber	.66	.65	.67	.65	.69 (.60, .77)	.67 (.57, .75)

[a] All correlation coefficients were significant (*P*=.00-.04), except for energy-adjusted sodium (*P*=.06) and deattenuated sodium assessed with Meal-Q as well as crude riboflavin (*P*=.06), energy-adjusted sodium (*P*=.12) and deattenuated sodium assessed with MiniMeal-Q.

[b] n=162 because of exclusion of 1 participant with implausibly high intake.

[c] n=160 because of exclusion of 3 participants with implausibly high intakes.

Figure 3. Bland-Altman plots with the weighed food record (WFR), Meal-Q, and MiniMeal-Q for (a) folate, (b) thiamine, (c) iron, and (d) selenium (N=163). Differences in intake between the WFR and the questionnaires are plotted against the mean of the 2 methods. The solid line indicates the reference line of zero difference. The long-dashed line shows the mean difference. The short-dashed lines show the 95% limits of agreement (mean difference ±2 SD). *: Energy-adjusted.



Reproducibility

Table 6 shows the absolute intake of micronutrients and fiber assessed with the 2 administered Meal-Q questionnaires in group 2. There were no statistically significant differences between the questionnaires. The proportion of participants classified into the same or adjacent quartile was 86% to 97% and in the extreme quartile 0% to 3% (Table 7). Crude ICCs were in the range of .45-.85 and energy-adjusted ICCs in the range of .50-.76 (Table 7).

Table 6. The median and interquartile range (IQR) of daily micronutrient and fiber intake assessed with Meal-Q 1 and Meal-Q 2 from group 2, and the median (IQR) difference in intake between the questionnaires (n=87).

Nutrients	Meal-Q 1		Meal-Q 2[a]		Difference[b] (Meal-Q 1–Meal-Q 2)	
	Median	IQR	Median	IQR	Median	IQR
Beta-carotene (µg)	3246	2776	2441	3626	126	1271
Thiamine (mg)	1.4	0.7	1.5	0.9	0.01	0.48
Riboflavin (mg)	1.7	0.7	1.6	0.7	0.03	0.69
Niacin (mg)	15	9	16	9	–0.06	4.96
Vitamin B_6 (mg)	2.0	1.0	1.9	1.0	0.03	0.67
Folate (µg)	320	187	332	218	14	105
Vitamin B_{12} (µg)	3.7	2.1	4.0	2.5	–0.18	1.44
Vitamin C (mg)	108	81	99	93	–0.78	41.10
Vitamin D (µg)	4.9	2.8	5.2	3.6	–0.19	3.32
Vitamin E (mg)	9.6	5.2	9.5	5.6	0.27	3.47
Calcium (mg)	860	449	888	454	8.95	299.27
Iron (mg)	13	8	13	10	–0.07	4.37
Magnesium (mg)	406	212	418	233	10	112
Phosphorus (mg)	1419	668	1483	518	1.92	430.63
Potassium (mg)	3208	1719	3116	1584	38	963
Selenium (µg)	43	23	46	23	–0.12	15.93
Sodium (mg)	2466	984	2499	1294	–75	757
Zinc (mg)	11.0	4.9	11.0	3.7	0.004	3.556
Fiber (g)	28	19	25	22	0.20	8.53

[a] Missing values on Meal-Q 2 for 4 participants.

[b] None were statistically significant using Wilcoxon signed rank test (*P*=.07-.96).

Table 7. Quartile cross-classifications of Meal-Q 1 and Meal-Q 2[a] from group 2, and crude and energy-adjusted intraclass correlation coefficients (ICC) (n=87).

Nutrients	Participants, %				ICC (95% CI)	
	Same quartile	Adjacent quartile	Same/adjacent quartile	Extreme quartile	Crude	Energy-adjusted
Beta-carotene (µg)	53	43	96	1	.85 (.79, .91)	.75 (.66, .84)
Thiamine (mg)	57	34	91	2	.54 (.40, .69)	.64 (.51, .76)
Riboflavin (mg)	59	33	92	3	.45 (.28, .62)	.63 (.51, .76)
Niacin (mg)	52	41	93	0	.66 (.54, .78)	.76 (.67, .85)
Vitamin B_6 (mg)	53	36	89	1	.49 (.33, .65)	.50 (.34, .66)
Folate (µg)	59	38	97	1	.71 (.60, .81)	.73 (.63, .83)
Vitamin B_{12} (µg)	59	31	90	1	.60 (.47, .74)	.65 (.53, .78)
Vitamin C (mg)	52	43	95	1	.80 (.73, .88)	.74 (.64, .83)
Vitamin D (µg)	43	43	86	3	.56 (.42, .70)	.55 (.41, .70)
Vitamin E (mg)	57	34	91	0	.73 (.64, .83)	.73 (.63, .83)
Calcium (mg)	51	38	89	1	.49 (.33, .65)	.66 (.54, .78)
Iron (mg)	51	41	92	2	.61 (.47, .74)	.61 (.48, .74)
Magnesium (mg)	66	31	97	1	.64 (.51, .76)	.73 (.64, .83)
Phosphorus (mg)	54	33	87	3	.46 (.29, .62)	.62 (.49, .75)
Potassium (mg)	56	38	94	0	.65 (.52, .77)	.80 (.73, .88)
Selenium (µg)	61	33	94	1	.64 (.52, .77)	.72 (.61, .82)
Sodium (mg)	57	38	95	1	.53 (.38, .68)	.59 (.45, .72)
Zinc (mg)	46	40	86	1	.50 (.35, .66)	.63 (.50, .76)
Fiber (g)	55	39	94	0	.77 (.69, .86)	.71 (.61, .82)

[a] Missing values on Meal-Q 2 for 4 participants.

Discussion

Principal Results

This validation study suggests Meal-Q and MiniMeal-Q are useful tools for ranking micronutrient and fiber intake in epidemiological studies, with the exception of sodium. Furthermore, Meal-Q's reproducibility results indicate good assessment reliability.

Regarding assessment of absolute intake, both questionnaires underestimated intake of most micronutrients as compared to the WFR. This underestimation may be partly explained by the methodological differences between the methods. A food record has an open-ended design and is aimed to assess the whole diet during a consecutive number of days. In contrast, a questionnaire has predefined items and frequencies and naturally cannot assess the entire diet. Rather, the aim of a questionnaire is to assess dietary intake in a way that enables ranking of low to high consumers. Because risk comparisons in epidemiological studies commonly are made between different strata of intake, the ranking ability of dietary intake is usually of more interest than assessment of absolute intake [14,20]. Therefore, we conclude that Meal-Q and MiniMeal-Q are useful instruments in an epidemiological setting.

The captured between-person variance in intake assessed with MiniMeal-Q as compared to Meal-Q demonstrated only a minor loss of information when using MiniMeal-Q despite having approximately 30% fewer food items. This indicates MiniMeal-Q is a valuable alternative when a shorter questionnaire is desirable.

Acknowledging that the evaluation of MiniMeal-Q is made with truncated Meal-Q data, comparisons between them should be interpreted carefully. Comparing our results to other validation studies should also be done with caution given that differences in study design and participant demographics may affect the results. Yet, bearing its limits in mind, such comparisons, which are commonly made, are crucial in evaluating a questionnaire's performance.

Comparison With Prior Work

The cross-classifications with the WFR showed both questionnaires to yield ranking agreements comparable to or better than other similar validation studies [21-25], of which 2 evaluated Web-based FFQs. The highest-ranking agreement for Meal-Q and MiniMeal-Q was seen for fiber with 89% to 90% placed into the same or adjacent quartile, which is greater than in some other studies [21,23-24]. The lowest ranking agreement was seen for sodium, as has been shown previously [21,23], and which likely reflects the difficulty in assessing salt intake.

The Bland-Altman plots showed that Meal-Q and MiniMeal-Q had difficulties in precision as seen by the large variance. This varying bias over the intake range was also indicated by the limits of agreement, which for some nutrients deviated from 5%. For most nutrients, the questionnaires did not perform as well in assessing high intakes. This might be explained by a limitation of food items, excessive grouping of several food items on each row, lack of high frequency alternatives, or the use of standard portion sizes for many food items. The overall large variance seen for most nutrients could arise from various sources (eg, a limited frequency range of the questionnaires and/or a high between-person variation in the WFR). Although the Bland-Altman method has been recommended for use in validation studies, it should be noted that we would not expect an absolute agreement between the questionnaires and the WFR because of their inherent methodological differences. Instead, the plots are helpful in assessing the magnitude of the inaccuracy and detecting potential varying bias. Despite the varying bias over the intake range seen in the Bland-Altman plots, the cross-classifications of quartiles indicated that both Meal-Q and MiniMeal-Q were able to yield a good ranking ability.

The limited number of studies using Bland-Altman plots for assessment of micronutrient validity and that some of them used log-transformed values makes comparisons with our results difficult. However, 2 other studies have also detected varying bias over the intake range. Labonté et al [23] showed similar results in variance for fiber intake and Pinto et al [25] showed a larger variance for folate and iron intake than seen in the present study.

The energy-adjusted and deattenuated Spearman correlation coefficients in the current study were similar to or better than correlations obtained in other validation studies with comparable study design [21-27]. Sodium showed a statistically nonsignificant correlation with the WFR for both questionnaires, which has also been seen previously [21,23]. Furthermore, in a review of 392 validation studies of vitamin intake, Henríquez-Sánchez et al [28] showed mean correlations between a FFQ and a dietary record in the range *r*=.41-.53. Another review of 109 validation studies of iron, calcium, selenium, and zinc reported mean correlations between a FFQ and a dietary record ranging from *r*=.36-.60 [29]. Both reviews show that the correlations in our study are in-line with other validation studies for most nutrients, with the exception of thiamine, riboflavin, vitamin B_6, vitamin D, vitamin E, calcium, and zinc, which had correlations somewhat lower in the present study. Only correlations for vitamin B_{12}, niacin, riboflavin, vitamin E, calcium, magnesium, selenium, and fiber improved after energy adjustment, a phenomenon also seen previously [25,26]. A possible explanation for this is varying correlation with energy between different nutrients [14], a feature that also depends on the population.

The use of correlation coefficients in validation studies is extensive, but has been criticized because they only measure a relationship and not the agreement between 2 methods [15]. However, as mentioned, we would not expect an absolute agreement between a FFQ and a food record because FFQs are designed to rank individuals rather than to assess absolute intake [20]. In this way, the correlation coefficient is a useful measure of validity because it assesses the ranking ability.

The sodium intake assessed with Meal-Q and MiniMeal-Q only included salt in food items and dishes in the nutrient database. Both questionnaires have a yes/no question regarding salt in cooking and table salt; however, because it is difficult to estimate amounts, this information was not included in the nutrient calculations. The WFR could potentially capture added salt; however, this was only reported for a minor fraction of all food items. Hence, the sodium assessed with the questionnaires and the WFR both originate from salt already present in food items and dishes from the nutrient database. Therefore, the low validity for sodium could best be explained by a general large random variation in assessment between the questionnaire and the WFR.

The reproducibility of Meal-Q indicated that it performed well in its reliability to rank dietary intake, with a high proportion of participants in the same/adjacent quartile and a low proportion of misclassified participants. The quartile cross-classifications were comparable to Labonté et al [23]. Energy-adjusted correlations between repeated FFQs have generally ranged between *r*=.5-.8 in other studies [30] and Meal-Q showed quite similar results. The ICCs for fiber, vitamin E, vitamin B_6, niacin, vitamin C, beta-carotene, folate, magnesium, and potassium in the present study were, on average, lower than those found by Schröder et al [31]. Furthermore, the ICCs were slightly lower than the Pearson correlations found by Labonté et al [23], yet higher than the Pearson correlations found by Pinto et al [25].

Limitations and Strengths

A strength of this study was the large sample size for this type of validation study. Moreover, there was low dropout and high compliance for the assessment methods throughout the entire study. The high compliance probably reflects a well-motivated study population, something that is vital for the study's internal validity. The motivation might arise from a general higher interest in health among self-selected participants as compared to invited participants. Furthermore, participants with nutritional backgrounds might also be more motivated than those without this background. It should be acknowledged that the young and primarily female study population might have implications on external validity. Regarding data handling, the Web-based format of the questionnaires and the WFR minimized potential errors in the conversion of crude consumption data into nutrient intakes. Web-based formats have previously shown to improve data quality [4,32].

In the validation of a dietary assessment method, the reference method should have measurement errors independent from those of the test method. Because the WFR is an open-ended prospective method and an FFQ is a retrospective method with predefined food items and frequencies, dependent measurement errors are less likely to occur. Nevertheless, both methods are susceptible to social desirability, as are all dietary assessment methods. This could affect them in similar ways and increase their error dependency as a result. Also, both methods are linked to the same nutrient database. Therefore, a validation study of a dietary assessment method should be evaluated for relative validity rather than absolute validity. Unfortunately, the present study did not have the means to include an objective reference method for micronutrient intake as biomarkers (eg, urinary potassium, thiamine, and sodium), which would have been a valuable complement to the WFR.

Meal-Q and MiniMeal-Q reflect dietary intake during the past few months, whereas the WFR captures dietary intake over 7 consecutive days; hence, a perfect agreement should not be expected. Ideally, the WFR would have been performed repeatedly over a longer time period to better mirror the assessment aim of the questionnaires. Also, for the reproducibility analysis, the second Meal-Q should have been administered after a slightly longer time period preferably to decrease the influence from the first questionnaire. However, time constraints made a longer validation study impossible. In the comparisons between questionnaires and the WFR, adjustments for within-person variance in the WFR were made to minimize the effect of day-to-day variations in intake. Furthermore, MiniMeal-Q ideally should have been evaluated in a separate validation study; however, this was not possible because of time constraints.

Conclusions

This validation study demonstrated that Meal-Q and MiniMeal-Q are useful questionnaires for ranking micronutrient and fiber intake in epidemiological studies using Web-based data collection. However, assessment of sodium intake requires further attention in future questionnaire versions. Furthermore, the reproducibility results showed Meal-Q to have good reliability. It should be noted that the study was conducted in a young, primarily female, and well-educated study population, and that MiniMeal-Q merits its own validation study.

Acknowledgments

We would like to thank the participants in the validation study. This work was supported by funds from Torsten and Ragnar Söderberg’s Foundation, AFA Insurance, the Swedish Research Council, and the Swedish Council for Working Life and Social Research.

Authors' Contributions

The authors’ responsibilities were questionnaire design: SEC, EM, SEB, LL, OB, KB; validation study design: SEC, EM, SEB, LL, KB; data collection: SEC, EM, SEB, KB; development and validation of the nutrient calculation programs MealCalc and MiniMealCalc: SEC, OB, KB; calculation of nutrients: OB; statistical analyses: SEC, AP; interpretation of results: SEC, AP, LL, KB; drafted the manuscript: SEC; review and revision of the manuscript including approval of final version: SEC, EM, SEB, AP, OB, LL, KB.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Bland-Altman plots with the WFR, Meal-Q and MiniMeal-Q for (e) beta-carotene (n=162 due to exclusion of one subject with implausibly high intake), (f) riboflavin, (g) niacin and (h) vitamin b6 (n=163). Differences in intake between the WFR and the questionnaires are plotted against the mean of the two methods. The solid line indicates the reference line of zero difference. The long-dashed line shows the mean difference. The short-dashed lines show the 95% limits of agreement (mean difference ±2 SD). * Energy-adjusted.

[JPG File, 141KB - jmir_v16i2e59_app1.jpg]

Multimedia Appendix 2

Bland-Altman plots with the WFR, Meal-Q and MiniMeal-Q for (i) vitamin B12, (j) vitamin C, (k) vitamin D and (l) vitamin E (n=163). Differences in intake between the WFR and the questionnaires are plotted against the mean of the two methods. The solid line indicates the reference line of zero difference.The long-dashed line shows the mean difference. The short-dashed lines show the 95% limits of agreement (mean difference ±2 SD). * Energy-adjusted.

[JPG File, 130KB - jmir_v16i2e59_app2.jpg]

Multimedia Appendix 3

Bland-Altman plots with the WFR, Meal-Q and MiniMeal-Q for (m) calcium, (n) magnesium, (o) phosphorus and (p) potassium (n=163). Differences in intake between the WFR and the questionnaires are plotted against the mean of the two methods. The solid line indicates the reference line of zero difference. The long-dashed line shows the mean difference. The short-dashed lines show the 95% limits of agreement (mean difference ±2 SD). * Energy-adjusted.

[JPG File, 140KB - jmir_v16i2e59_app3.jpg]

Multimedia Appendix 4

Bland-Altman plots with the WFR, Meal-Q and MiniMeal-Q for (q) sodium (n=160 due to exclusion of three subjects with implausibly high intakes), (r) zinc and (s) fiber (n=163). Differences in intake between the WFR and the questionnaires are plotted against the mean of the two methods. The solid line indicates the reference line of zero difference. The long-dashed line shows the mean difference. The short-dashed lines show the 95% limits of agreement (mean difference ±2 SD). * Energy-adjusted.

[JPG File, 98KB - jmir_v16i2e59_app4.jpg]

References

1. International Telecommunication Union. 2013. Statistics URL: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2013-09-02] [WebCite Cache ID 6JKaoWUmJ]
2. Statistics Sweden. 2013. Use of computers and the Internet by private persons in 2012 URL: http://www.scb.se/statistik/_publikationer/LE0108_2012A01_BR_IT01BR1301.pdf [accessed 2013-09-02] [WebCite Cache ID 6JKb9OzHP]
3. Greenlaw C, Brown-Welty S. A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost. Eval Rev 2009 Oct;33(5):464-480. [doi: 10.1177/0193841X09340214] [Medline: 19605623]
4. Illner AK, Freisling H, Boeing H, Huybrechts I, Crispim SP, Slimani N. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. Int J Epidemiol 2012 Aug;41(4):1187-1203 [FREE Full text] [doi: 10.1093/ije/dys105] [Medline: 22933652]
5. Ekman A, Litton JE. New times, new needs; e-epidemiology. Eur J Epidemiol 2007;22(5):285-292. [doi: 10.1007/s10654-007-9119-0] [Medline: 17505896]
6. Bälter KA, Bälter O, Fondell E, Lagerros YT. Web-based and mailed questionnaires: a comparison of response rates and compliance. Epidemiology 2005 Jul;16(4):577-579. [Medline: 15951679]
7. Subar AF, Thompson FE, Potischman N, Forsyth BH, Buday R, Richards D, et al. Formative research of a quick list for an automated self-administered 24-hour dietary recall. J Am Diet Assoc 2007 Jun;107(6):1002-1007. [doi: 10.1016/j.jada.2007.03.007] [Medline: 17524721]
8. Kohlmeier L. Gaps in dietary assessment methodology: meal- vs list-based methods. Am J Clin Nutr 1994 Jan;59(1 Suppl):175S-179S [FREE Full text] [Medline: 8279419]
9. Christensen SE, Möller E, Bonn SE, Ploner A, Wright A, Sjölander A, et al. Two new meal- and web-based interactive food frequency questionnaires: validation of energy and macronutrient intake. J Med Internet Res 2013;15(6):e109 [FREE Full text] [doi: 10.2196/jmir.2458] [Medline: 23739995]
10. Uppsala: National Food Agency. URL: http://www7.slv.se/Naringssok/?epslanguage=sv [accessed 2012-11-27] [WebCite Cache ID 6CU63QALA]
11. Davidsson L. Assessment of body composition and total energy expenditure in humans using stable isotope techniques. IAEA Human Health Series. Vienna: International Atomic Energy Agency; 2009. URL: http://www-pub.iaea.org/books/IAEABooks/7982/Assessment-of-Body-Composition-and-Total-Energy-Expenditure-in-Humans-Using-Stable-Isotope-Techniques [accessed 2013-09-02] [WebCite Cache ID 6JKdOJZ8S]
12. Bonn SE, Trolle Lagerros Y, Christensen SE, Möller E, Wright A, Sjölander A, et al. Active-Q: validation of the web-based physical activity questionnaire using doubly labeled water. J Med Internet Res 2012;14(1):e29 [FREE Full text] [doi: 10.2196/jmir.1974] [Medline: 22356755]
13. Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. Int J Obes Relat Metab Disord 2000 Sep;24(9):1119-1130. [Medline: 11033980]
14. Willett W. Nutritional Epidemiology, 2nd ed. New York: Oxford University Press; 1998.
15. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
16. Beaton GH, Milner J, McGuire V, Feather TE, Little JA. Source of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Carbohydrate sources, vitamins, and minerals. Am J Clin Nutr 1983 Jun;37(6):986-995 [FREE Full text] [Medline: 6846242]
17. Liu K, Stamler J, Dyer A, McKeever J, McKeever P. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. J Chronic Dis 1978;31(6-7):399-418. [Medline: 711832]
18. Rosner B, Willett WC. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. Am J Epidemiol 1988 Feb;127(2):377-386. [Medline: 3337089]
19. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. Comput Biol Med 1989;19(1):61-70. [Medline: 2917462]
20. Masson LF, McNeill G, Tomany JO, Simpson JA, Peace HS, Wei L, et al. Statistical approaches for assessing the relative validity of a food-frequency questionnaire: use of correlation coefficients and the kappa statistic. Public Health Nutr 2003 May;6(3):313-321. [doi: 10.1079/PHN2002429] [Medline: 12740081]

21. Chacko George G, Milani TJ, Hanss-Nuss H, Kim M, Freeland-Graves JH. Development and validation of a semi-quantitative food frequency questionnaire for young adult women in the southwestern United States. Nutrition Research 2004 Jan;24(1):29-43. [doi: 10.1016/j.nutres.2003.09.006]
22. Heath AL, Skeaff CM, Gibson RS. The relative validity of a computerized food frequency questionnaire for estimating intake of dietary iron and its absorption modifiers. Eur J Clin Nutr 2000 Jul;54(7):592-599. [Medline: 10918471]
23. Labonté M, Cyr A, Baril-Gravel L, Royer MM, Lamarche B. Validity and reproducibility of a web-based, self-administered food frequency questionnaire. Eur J Clin Nutr 2012 Feb;66(2):166-173. [doi: 10.1038/ejcn.2011.163] [Medline: 21934698]
24. Pakseresht M, Sharma S, Cao X, Harris R, Caberto C, Wilkens LR, Barbados National Cancer Study Group. Validation of a quantitative FFQ for the Barbados National Cancer Study. Public Health Nutr 2011 Mar;14(3):426-434. [doi: 10.1017/S1368980010002351] [Medline: 20843400]
25. Pinto E, Severo M, Correia S, dos Santos Silva I, Lopes C, Barros H. Validity and reproducibility of a semi-quantitative food frequency questionnaire for use among Portuguese pregnant women. Matern Child Nutr 2010 Apr;6(2):105-119. [doi: 10.1111/j.1740-8709.2009.00199.x] [Medline: 20624208]
26. Beasley JM, Davis A, Riley WT. Evaluation of a web-based, pictorial diet history questionnaire. Public Health Nutr 2009 May;12(5):651-659 [FREE Full text] [doi: 10.1017/S1368980008002668] [Medline: 18547450]
27. Lassale C, Guilbert C, Keogh J, Syrette J, Lange K, Cox DN. Estimating food intakes in Australia: validation of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) food frequency questionnaire against weighed dietary intakes. J Hum Nutr Diet 2009 Dec;22(6):559-566. [doi: 10.1111/j.1365-277X.2009.00990.x] [Medline: 19735348]
28. Henríquez-Sánchez P, Sánchez-Villegas A, Doreste-Alonso J, Ortiz-Andrellucchi A, Pfrimer K, Serra-Majem L. Dietary assessment methods for micronutrient intake: a systematic review on vitamins. Br J Nutr 2009 Dec;102 Suppl 1:S10-S37. [doi: 10.1017/S0007114509993126] [Medline: 20100364]
29. Serra-Majem L, Pfrimer K, Doreste-Alonso J, Ribas-Barba L, Sánchez-Villegas A, Ortiz-Andrellucchi A, et al. Dietary assessment methods for intakes of iron, calcium, selenium, zinc and iodine. Br J Nutr 2009 Dec;102 Suppl 1:S38-S55. [doi: 10.1017/S0007114509993138] [Medline: 20100367]
30. Gibson RS. Principles of Nutritional Assessment, 2nd ed. New York: OUP Australia and New Zealand; 2005.
31. Schröder H, Covas MI, Marrugat J, Vila J, Pena A, Alcántara M, et al. Use of a three-day estimated food record, a 72-hour recall and a food-frequency questionnaire for dietary assessment in a Mediterranean Spanish population. Clin Nutr 2001 Oct;20(5):429-437. [doi: 10.1054/clnu.2001.0460] [Medline: 11534938]
32. van Gelder MM, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol 2010 Dec 1;172(11):1292-1298 [FREE Full text] [doi: 10.1093/aje/kwq291] [Medline: 20880962]

Abbreviations

BMI: body mass index
DLW: doubly labeled water
FFQ: food frequency questionnaire
ICC: intraclass correlation coefficient
PAL: physical activity level
WFR: weighed food record

Edited by G Eysenbach; submitted 17.09.13; peer-reviewed by AL Brantsaeter, E Barrat; comments to author 23.12.13; revised version received 02.01.14; accepted 12.01.14; published 21.02.14.

Please cite as:
Christensen SE, Möller E, Bonn SE, Ploner A, Bälter O, Lissner L, Bälter K
Relative Validity of Micronutrient and Fiber Intake Assessed With Two New Interactive Meal- and Web-Based Food Frequency Questionnaires
J Med Internet Res 2014;16(2):e59
URL: http://www.jmir.org/2014/2/e59/
doi:10.2196/jmir.2965
PMID:24565605

Original Paper

Effect of a Kinect-Based Exercise Game on Improving Executive Cognitive Performance in Community-Dwelling Elderly: Case Control Study

Hiroki Kayama[1], PT; Kazuya Okamoto[2], PhD; Shu Nishiguchi[1], PT; Minoru Yamada[1], PT, PhD; Tomohiro Kuroda[2], PhD; Tomoki Aoyama[1], MD, PhD

[1]Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan

Corresponding Author:
Hiroki Kayama, PT
Human Health Sciences
Graduate School of Medicine
Kyoto University
53, Kawaharcho, Shogoin
Sakyo-ku
Kyoto, 606-8507
Japan
Phone: 81 75 751 3909
Fax: 81 75 751 3964
Email: kayama.hiroki.37r@st.kyoto-u.ac.jp

Abstract

Background: Decrease of dual-task (DT) ability is known to be one of the risk factors for falls. We developed a new game concept, Dual-Task Tai Chi (DTTC), using Microsoft's motion-capture device Kinect, and demonstrated that the DTTC test can quantitatively evaluate various functions that are known risk factors for falling in elderly adults. Moreover, DT training has been attracting attention as a way to improve balance and DT ability. However, only a few studies have reported that it improves cognitive performance.

Objective: The purpose of this study was to demonstrate whether or not a 12-week program of DTTC training would effectively improve cognitive functions.

Methods: This study examined cognitive functions in community-dwelling older adults before and after 12 weeks of DTTC training (training group [TG]) or standardized training (control group [CG]). Primary end points were based on the difference in cognitive functions between the TG and the CG. Cognitive functions were evaluated using the trail-making test (part A and part B) and verbal fluency test.

Results: A total of 41 elderly individuals (TG: n=26, CG: n=15) participated in this study and their cognitive functions were assessed before and after DTTC training. Significant differences were observed between the two groups with significant group × time interactions for the executive cognitive function measure, the delta-trail-making test (part B−part A; $F_{1,36}$=4.94, *P*=.03; TG: pre mean 48.8 [SD 43.9], post mean 42.2 [SD 29.0]; CG: pre mean 49.5 [SD 51.8], post mean 64.9 [SD 54.7]).

Conclusions: The results suggest that DTTC training is effective for improving executive cognitive functions.

Trial Registration: Japan Medical Association Clinical Trial Registration Number: JMA-IIA00092; https://dbcentre3.jmacct.med.or.jp/jmactr/App/JMACTRS06/JMACTRS06.aspx?seqno=2682 (Archived by WebCite at http://www.webcitation.org/6NRtOkZFh).

(J Med Internet Res 2014;16(2):e61) doi:10.2196/jmir.3108

KEYWORDS

fall prevention; cognitive function; dual-task; training; elderly

Introduction

Cognitive impairment among elderly individuals is a serious issue in many countries. Many investigators have developed different cognitive function training methods as countermeasures to prevent cognitive impairment and have reported their effects [1]. Other investigators also have reported the effects of physical exercise training on cognitive functions among elderly individuals [2,3]. Additionally, Hillman et al and Silsupado have indicated that executive cognitive functions, which are related to the control of goal-oriented actions and adaptive behaviors, are strongly impaired by aging and respond positively to exercise training [4,5].

Recently, dual-task (DT) ability, or the performance of simultaneous motor and cognitive tasks, has been receiving considerable attention [6]. DT training is now recognized as a fall prevention tool that enhances physical functions among elderly people [7].

With a focus on DT, we developed a new concept called the Dual-Task Tai Chi (DTTC) test [8]. This system was developed using Kinect (Microsoft, Redmond, WA, USA), a motion-capture device, and demonstrated that the DTTC test quantitatively evaluates compound functions, including DT, balance, and cognitive abilities in elderly people [9]. In unpublished data, we found that DTTC training was useful not only to assess but also to improve balance and mobility among elderly people [10].

Several investigators have reported the effects of DT training on balance, mobility, walking, and DT ability [11,12]. However, only a few have reported that DT training improves cognitive performance. We reported that Trail-Walking Exercise, which is similar to Trail-Making Test (TMT) under DT condition, improved executive cognitive functions [13]. According to this, we expected that DTTC training would improve cognitive functions as well, especially executive functions. Therefore, the purpose of this study was to reveal that training with the DTTC device affects cognitive performance in elderly individuals.

Methods

Participants

Community-dwelling elderly subjects (n=48) participated in this study. The subjects were recruited through an advertisement in the local press. The following selection criteria were used: age ≥65 years, community dwelling, independent ambulation, willingness to participate in the measurement of physical fitness, and minimal hearing and vision impairment. Exclusion criteria were as follows: inability to complete the tasks because of reduced cognitive functions, evaluated by Rapid Dementia Screening Test [14] scored 8 or greater; severe cardiac, pulmonary, or musculoskeletal disorders; pathologies associated with increased risk of falls, such as Parkinson’s disease or stroke; osteoporosis; and psychotropic drug use. We obtained written informed consent from each participant. This study was approved (protocol approval E- 880) by the Ethical Review Board of Kyoto University Graduate School of Medicine, Kyoto, Japan.

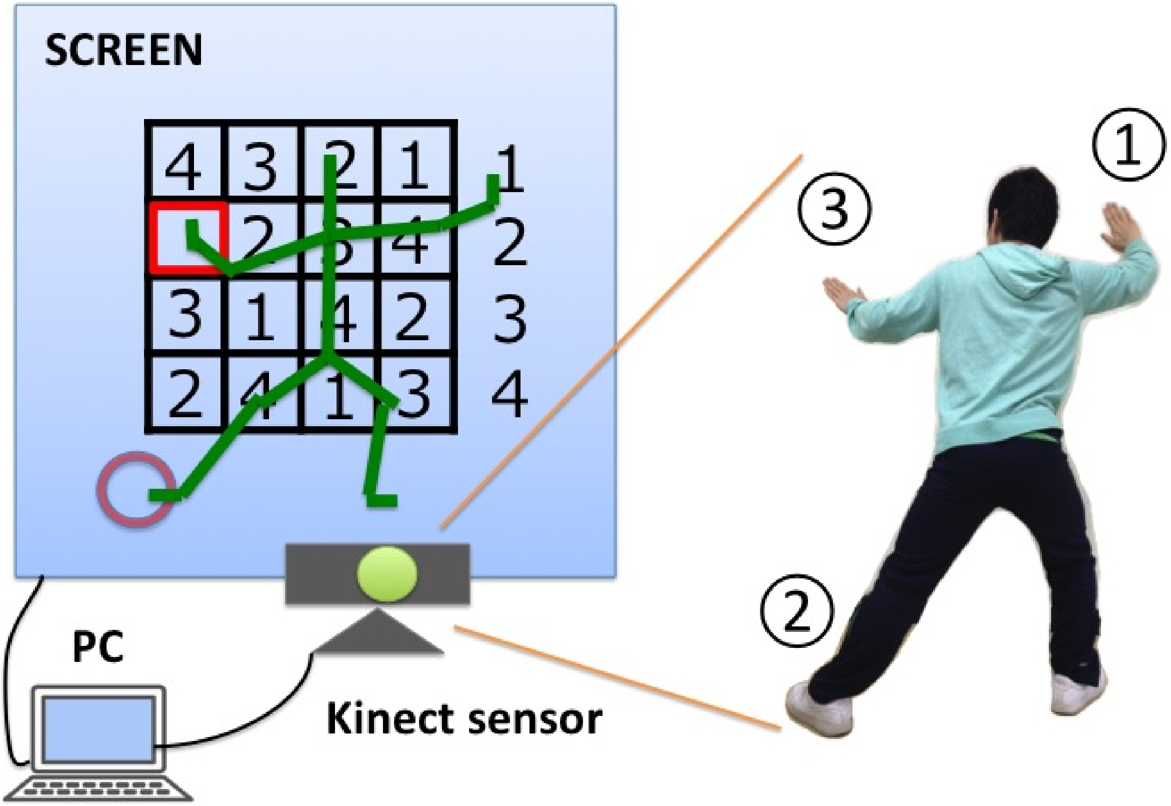
Device

The DTTC test [8-10] requires users to solve a number placement problem (Sudoku) by controlling a stick figure with the movement of their entire body. The user’s full-body motion is captured using Kinect and is translated into movements for the stick figure on a screen. The cognitive task is to fill in 3 boxes chosen at random from a 4 × 4 grid with digits ranging from 1 to 4. The user selects a digit using his or her right hand and left foot and points to a box with his or her left hand. In addition, the user must move his or her right hand to the left hand to fill the indicated box with the selected digit. As such, full-body motion, similar to Tai Chi Chuan movements, is required. We recorded the time taken to fill in all 3 boxes, as our evaluation index.

To begin with, the user stands 3 m in front of the Kinect sensor with his or her right foot in front of the sensor (Figure 1). The following instructions were provided:

1. Reach a digit you need to use with your right hand to fill a blank you want to answer.
2. Step 50 cm laterally, with your left leg, to grip the digit in your right hand.
3. Select the blank you want to answer with your left hand, and move your right hand to your left one.

Figure 1. View of the Dual-Task Tai Chi (DTTC) test.



Intervention

All subjects participated in group training sessions lasting 75-80 minutes once a week for 12 weeks. The participants were divided into two training groups according to their participation in the exercise class: (1) the control group (CG), 75-minute standardized training, and (2) the training group (TG), 5-minute DTTC training, in addition to the standardized training [10].

The exercise classes were individualized for each group and were supervised by physical therapists. Each exercise class used a standardized format that included 15 minutes of moderate-intensity aerobic exercise, 15 minutes of progressive strength training, 10 minutes of flexibility and balance exercises, and 10 minutes of cool-down activities, followed by exercises known to improve muscle strength and balance [15,16]. In addition, the class included a 25-minute rhythmic stepping exercise involving cognitive ability, which is an exercise intended to improve DT ability [11], before cool-down.

The participants in the TG were additionally asked to solve DTTC problems and mirror-reversed DTTC problems, alternately, as many times possible in 5 minutes.

Outcome Measures

All participants underwent evaluation upon entry into the study (pre-intervention) and at the end of the study (post-intervention), using the results of 2 cognitive performance tests.

Cognitive functions were evaluated using the trail-making test (TMT) [17] and verbal fluency test (VFT) [18-20]. The TMT is a well-established psychomotor test originally developed as part of the Army Individual Test Battery. The TMT has been widely used in clinical evaluations to assess deficits in executive cognitive functions. The test consists of 2 parts: part A is a visual-scanning task and part B is a measure of cognitive flexibility. For this analysis, we used a different score defined as delta-TMT, calculated as the difference between the times for each part (part B-part A). The delta-TMT score is used to control for the effect of motor speed on TMT performance and is considered a more accurate measure of executive functions than performance on part B alone [20,21].

The VFT has a letter fluency component and a category fluency component. Participants were asked to think of as many animal names as possible in 1 minute (category fluency). Verbal fluency is an evaluation of expressive language ability and executive functions. The score was the number of successful words (except for some proper nouns).

Statistical Analysis

We compared baseline characteristics between the participants in each group to examine the comparability between the 2 groups using Student's *t* test or the chi-square test. Repeated-measures, mixed-linear, two-way ANCOVA (analysis of covariance) was used to analyze the effect of exercise on outcome measurements while adjusting for each cognitive performance, at pre-intervention, as a covariate.

Data were entered and analyzed using the Statistical Package for the Social Sciences (Windows version 20.0, SPSS Inc., Chicago, IL, USA). For all analyses, $P<.05$ was considered statistically significant.

Results

Study Population

A total of 41 of the 48 selected subjects (85.4%) completed the study protocols and returned for their exit interviews and final testing (TG: n=26, CG: n=15). The participants' baseline data did not differ significantly between the two groups. Thus, the groups were comparable and well matched with regard to their baseline characteristics.

Adherence to the Study Protocols

During the 12-week intervention phase, 10 exercise sessions were scheduled and all took place. Excluding the 4 participants who dropped out, the TG subjects had an overall attendance rate of 82% and the CG subjects had an overall attendance rate of 81% over the 12 weeks. No health problems, including cardiovascular or musculoskeletal complications, occurred during training sessions or testing. Moreover, almost all participants seemed to have enjoyed the DTTC training. They shared many positive opinions after each session and seemed to look forward to playing DTTC once a week.

Evaluation Outcomes

Pre- and post-intervention group statistics and group × time interactions are shown in Table 1. There was a significant difference between the groups regarding the changes (intervention to baseline) in delta-TMT. There were no significant differences among the other measures.

Table 1. Outcome measures by group at pre- and post-intervention.

Measures		Pre-intervention, mean (SD)	Post-intervention, mean (SD)	Group × Time *F* value	Degrees of freedom	*P* value
VFT [a]						
	TG[b]	11.96 (3.55)	12.04 (3.26)	0.09	1,38	.76
	CG[c]	11.38 (4.21)	11.38 (4.07)			
TMT-A [d]						
	TG	71.6 (23.8)	68.4 (19.9)	0.51	1,38	.48
	CG	82.2 (27.1)	70.0 (15.5)			
TMT-B						
	TG	120.3 (55.1)	110.4 (39.2)	2.73	1,38	.11
	CG	131.8 (62.6)	134.9 (61.1)			
ΔTMT						
	TG	48.8 (43.9)	42.2 (29.0)	4.94	1,36	.03
	CG	49.5 (51.8)	64.9 (54.7)			

[a] VFT: verbal fluency test

[b] TG: training group

[c] CG: control group

[d] TMT: trail-making test

Discussion

Principal Findings

The delta-TMT score was significantly improved after DTTC training. The results suggested that DTTC training was effective in improving executive cognitive functions. In our unpublished data, DTTC training was useful for improving balance ability and mobility among elderly people [10]. Thus, DTTC training has the capability of improving both physical and cognitive functions. Executive cognitive functions are closely related to DT performance and are good predictors of falling [20]. Thus, an improvement in executive functions, by DTTC training, has a positive impact on DT ability and the prevention of falling.

Conversely, the TMT-A and -B scores were not significantly improved. In the TMT-A, both groups improved similarly. On the other hand, in the TMT-B, only the TG had a tendency to improve the score, while CG showed little change. That is why the score of delta-TMT in CG increased and in TG decreased. The TMT-A and -B are used to assess visual scanning, cognitive flexibility, and executive functions [17]. The delta-TMT score is considered a more accurate measurement of executive functions [21,22]. That is, the results reflect specific improvement in executive cognitive functions. The reason is that DTTC training involves executive tasks that are the control of goal-oriented actions and adaptive behaviors. Additionally, VFT was also not significantly improved after DTTC training. It includes recalling tasks as their main elements; however, DTTC training does not involve recalling the task. This is considered the reason why the result of VFT was not significantly changed.

Previous studies reported that other “exergames,” based on Nintendo Wii Fit (Nintendo, Kyoto, Japan) or other devices, improved balance and leg muscle functions [23,24]. However, in these cases, the user needs control devices to capture motions and longer and more frequent exercise is required, compared with DTTC training, to obtain physical effects. DTTC overcomes these weaknesses and can offer comparable cognitive and physical benefits to users. Other previous studies have shown that a home-based dance device, using videogame technology, also improved the physical and cognitive parameters of fall risk in elderly people [25]. These home-based training tools are expected to increase the improvement of various functions and lead to fall prevention; our DTTC device is similar to these tools.

Limitations

There are several limitations to this study. First, the intervention effects in this study were not due solely to the DTTC training. That means the rhythmic stepping exercise involving cognitive ability in both groups also improves cognitive ability. Second, participants in both groups may have had higher motivation and interest in health issues and fall risk minimization than the general elderly population. An investigation into the effects of DTTC training on functions in frail, elderly adults is necessary, in the future.

Conclusions

In this study, the ANCOVA results of delta-TMT reveal that group × time interactions were statistically significant. They suggest that DTTC training is effective at improving executive cognitive functions in particular.

Acknowledgments

The authors thank the participants and are grateful to the department staff for helpful advice.

Conflicts of Interest

None declared.

References

1. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: a systematic review. PLoS One 2012;7(7):e40588 [FREE Full text] [doi: 10.1371/journal.pone.0040588] [Medline: 22792378]
2. Kamegaya T, Long-Term-Care Prevention Team of Maebashi City, Maki Y, Yamagami T, Yamaguchi T, Murai T, et al. Pleasant physical exercise program for prevention of cognitive decline in community-dwelling elderly with subjective memory complaints. Geriatr Gerontol Int 2012 Oct;12(4):673-679. [doi: 10.1111/j.1447-0594.2012.00840.x] [Medline: 22469534]
3. Kramer A, Hahn S, Cohen N, Banich M, McAuley E, Harrison C, et al. Ageing, Fitness and Neurocognitive Function. Nature 1999;400:418-419. [doi: 10.1038/22682]
4. Hillman CH, Erickson KI, Kramer AF. Be smart, exercise your heart: exercise effects on brain and cognition. Nat Rev Neurosci 2008 Jan;9(1):58-65. [doi: 10.1038/nrn2298] [Medline: 18094706]
5. Silsupadol P, Shumway-Cook A, Lugade V, van Donkelaar P, Chou LS, Mayr U, et al. Effects of single-task versus dual-task training on balance performance in older adults: a double-blind, randomized controlled trial. Arch Phys Med Rehabil 2009 Mar;90(3):381-387 [FREE Full text] [doi: 10.1016/j.apmr.2008.09.559] [Medline: 19254600]
6. Lundin-Olsson L, Nyberg L, Gustafson Y. Attention, frailty, and falls: the effect of a manual task on basic mobility. J Am Geriatr Soc 1998 Jun;46(6):758-761. [Medline: 9625194]
7. Yamada M, Aoyama T, Arai H, Nagai K, Tanaka B, Uemura K, et al. Dual-task walk is a reliable predictor of falls in robust elderly adults. J Am Geriatr Soc 2011 Jan;59(1):163-164 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.03206.x] [Medline: 21226689]
8. Okamoto H, Kayama M, Yamada M. The development of a system for the assessment of a dual-task performance based on a motion-capture device. 2012 Presented at: 9th International Conference on Disability, Virtual Reality & Associated Technologies; Sept 10 - 12, 2012; Laval, France p. 483.
9. Kayama H, Okamoto K, Nishiguchi S, Nagai K, Yamada M, Aoyama T. Concept software based on Kinect for assessing dual-task ability of elderly people. Games for Health Journal 2012 Oct;1(5):348-352. [doi: 10.1089/g4h.2012.0019]
10. Kayama H, Okamoto K, Nishiguchi S, Yukutake T, Tanigawa T, Nagai K, et al. Efficacy of an exercise game based on Kinect in improving physical performances of fall risk factors in community-dwelling older adults. Games for Health Journal 2013 Aug;2(4):247-252. [doi: 10.1089/g4h.2013.0006]
11. Yamada M, Tanaka B, Nagai K, Aoyama T, Ichihashi N. Rhythmic stepping exercise under cognitive conditions improves fall risk factors in community-dwelling older adults: Preliminary results of a cluster-randomized controlled trial. Aging Ment Health 2011 Jul 1;15(5):647-653. [doi: 10.1080/13607863.2010.551341] [Medline: 21815857]

12. Melzer I, Oddsson L. Improving balance control and self-reported lower extremity function in community-dwelling older adults: a randomized control trial. Clin Rehabil 2013 Mar;27(3):195-206. [doi: 10.1177/0269215512450295] [Medline: 22837541]
13. Yamada M, Tanaka B, Nagai K, Aoyama T, Ichihashi N. Trail-walking exercise and fall risk factors in community-dwelling older adults: preliminary results of a randomized controlled trial. J Am Geriatr Soc 2010 Oct;58(10):1946-1951. [doi: 10.1111/j.1532-5415.2010.03059.x] [Medline: 20831723]
14. Kalbe E, Calabrese P, Schwalen S, Kessler J. The Rapid Dementia Screening Test (RDST): a new economical tool for detecting possible patients with dementia. Dement Geriatr Cogn Disord 2003;16(4):193-199. [Medline: 14512713]
15. Buchner DM, Cress ME, de Lateur BJ, Esselman PC, Margherita AJ, Price R, et al. The effect of strength and endurance training on gait, balance, fall risk, and health services use in community-living older adults. J Gerontol A Biol Sci Med Sci 1997 Jul;52(4):M218-M224. [Medline: 9224433]
16. Gardner M, Robertson M, Campbell A. Exercise in preventing falls and fall related injuries in older people: a review of randomised controlled trials. Br J Sports Med 2000 Feb;34(1):7-17. [doi: 10.1136/bjsm.34.1.7]
17. War Department Adjutant General's Office. Manual of directions and scoring. In: Army Individual Test Battery. Washington DC: War Department; 1944.
18. Benton AL. Differential behavioral effects in frontal lobe disease. Neuropsychologia 1968 Mar;6(1):53-60. [doi: 10.1016/0028-3932(68)90038-9]
19. Nemoto T, Kashima H, Mizuno M. Contribution of divergent thinking to community functioning in schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry 2007 Mar 30;31(2):517-524. [doi: 10.1016/j.pnpbp.2006.12.001] [Medline: 17218048]
20. Mirelman A, Herman T, Brozgol M, Dorfman M, Sprecher E, Schweiger A, et al. Executive function and falls in older adults: new findings from a five-year prospective study link fall risk to cognition. PLoS One 2012;7(6):e40297 [FREE Full text] [doi: 10.1371/journal.pone.0040297] [Medline: 22768271]
21. Lezak MD, Howieson DB, Bigler ED, Tranel DD. Neuropsychological Assessment. New York: Oxford University Press, USA; 2004.
22. Corrigan JD, Hinkeldey NS. Relationships between parts A and B of the Trail Making Test. J Clin Psychol 1987 Jul;43(4):402-409. [Medline: 3611374]
23. Jorgensen MG, Laessoe U, Hendriksen C, Nielsen OB, Aagaard P. Efficacy of Nintendo Wii training on mechanical leg muscle function and postural balance in community-dwelling older adults: a randomized controlled trial. J Gerontol A Biol Sci Med Sci 2013 Jul;68(7):845-852. [doi: 10.1093/gerona/gls222] [Medline: 23114461]
24. Sullivan B, Harding AG, Dingley J, Gras LZ. Improvements in dynamic balance using an adaptive snowboard with the Nintendo Wii. Games for Health Journal 2012 Aug;1(4):269-273. [doi: 10.1089/g4h.2012.0006]
25. Schoene D, Lord SR, Delbaere K, Severino C, Davies TA, Smith ST. A randomized controlled pilot study of home-based step training in older people using videogame technology. PLoS One 2013;8(3):e57734 [FREE Full text] [doi: 10.1371/journal.pone.0057734] [Medline: 23472104]

Abbreviations

CG: control group
DT: dual task
DTTC: Dual-Task Tai Chi
TG: training group
TMT: trail-making test
VFT: verbal fluency test

Edited by H Fardoun; submitted 19.11.13; peer-reviewed by V Penichet, G Wolbring; comments to author 09.12.13; revised version received 15.01.14; accepted 22.01.14; published 24.02.14.

Please cite as:
Kayama H, Okamoto K, Nishiguchi S, Yamada M, Kuroda T, Aoyama T
Effect of a Kinect-Based Exercise Game on Improving Executive Cognitive Performance in Community-Dwelling Elderly: Case Control Study
J Med Internet Res 2014;16(2):e61
URL: http://www.jmir.org/2014/2/e61/
doi:10.2196/jmir.3108
PMID:24565806

Original Paper

The 1% Rule in Four Digital Health Social Networks: An Observational Study

Trevor van Mierlo[1,2], BA(Hons), MScCH, MBA, GEMBA HSG

[1]Research Associate, Henley Business School, University of Reading, Greenlands, Henley-on-Thames, United Kingdom
[2]Evolution Health Systems Inc, Toronto, ON, Canada

Corresponding Author:
Trevor van Mierlo, BA(Hons), MScCH, MBA, GEMBA HSG
Evolution Health Systems Inc
1266 Queen Street West
Suite 8
Toronto, ON, M6K 1L3
Canada
Phone: 1 416 644 8476 ext 221
Fax: 1 416 644 8477
Email: tvanmierlo@evolutionhs.com

Abstract

Background: In recent years, cyberculture has informally reported a phenomenon named the 1% rule, or 90-9-1 principle, which seeks to explain participatory patterns and network effects within Internet communities. The rule states that 90% of actors observe and do not participate, 9% contribute sparingly, and 1% of actors create the vast majority of new content. This 90%, 9%, and 1% are also known as Lurkers, Contributors, and Superusers, respectively. To date, very little empirical research has been conducted to verify the 1% rule.

Objective: The 1% rule is widely accepted in digital marketing. Our goal was to determine if the 1% rule applies to moderated Digital Health Social Networks (DHSNs) designed to facilitate behavior change.

Methods: To help gain insight into participatory patterns, descriptive data were extracted from four long-standing DHSNs: the AlcoholHelpCenter, DepressionCenter, PanicCenter, and StopSmokingCenter sites.

Results: During the study period, 63,990 actors created 578,349 posts. Less than 25% of actors made one or more posts. The applicability of the 1% rule was confirmed as Lurkers, Contributors, and Superusers accounted for a weighted average of 1.3% (n=4668), 24.0% (n=88,732), and 74.7% (n=276,034) of content.

Conclusions: The 1% rule was consistent across the four DHSNs. As social network sustainability requires fresh content and timely interactions, these results are important for organizations actively promoting and managing Internet communities. Superusers generate the vast majority of traffic and create value, so their recruitment and retention is imperative for long-term success. Although Lurkers may benefit from observing interactions between Superusers and Contributors, they generate limited or no network value. The results of this study indicate that DHSNs may be optimized to produce network effects, positive externalities, and bandwagon effects. Further research in the development and expansion of DHSNs is required.

(J Med Internet Res 2014;16(2):e33) doi:10.2196/jmir.2966

KEYWORDS

social networks; Superusers; eHealth; 1% rule; Pareto Principal; 90-9-1 principle; moderated support

Introduction

Background

Research examining digital health social networks (DHSNs) and their feasibility to improve health began in the mid-1980s [1,2]. As these networks became increasingly available, studies focused on relationships between network size, structure, program sustainability [3-5], and motivations of participants [6].

Terminology also developed to define common roles and behavior. For example, members of social networking sites (SNS) are now commonly referred to as actors [7]. Lurking (or passively reading social network conversations without actively participating) is the most common behavior [8]. Conversely, common to SNS are actors who frequently generate content and facilitate discussions [9-12]. In practice, these actors are often referred to as Superusers [13]. The impact or value that Lurkers, Superusers, or other actors have within DHSNs has not been empirically examined.

Network Effects and Positive Network Externalities

To frame their value, it may be beneficial to view Lurkers, Contributors, and Superusers and other actors through the lens of sociology, political science, economics, and finance where there is a rich history of examining network effects. A network effect occurs when an individual's use of a good or service influences its perceived value [14,15].

An example of a network effect can be seen in the popularity and growth of the fax machine. When few organizations had fax machines, the value of having a fax machine was low. However, as more organizations purchased fax machines and quickly and efficiently communicated with other departments or organizations, the network of fax machines grew and so did the value of owning one. Over time, having a fax machine in the workplace became essential. This is also known as the bandwagon effect, where the demand for a good increases because others are consuming it [16].

In the above example, the addition of each fax machine created a positive externality [17]. Positive externalities contribute to growth and popularity of a product or good, and social science research is now beginning to investigate this phenomenon within SNS [18].

Introducing the concept of network effects and positive externalities can help explain the importance of recruiting, retaining, and managing different types of actors to help grow DHSNs. If growing a DHSN increases program efficacy, it is important to understand the mechanisms behind content generation and how to increase network effects.

The 1% Rule (90-9-1 Principle)

Mirroring the well-established Pareto Principle, also known as the 80-20 rule [19], cyberculture and digital marketing have informally adopted a phenomenon named the 1% rule, or 90-9-1 principle [20,21]. Following the principals of a power law, the Pareto Principle is a natural observation illustrating that roughly 80% of effects come from 20% of causes [22]. Similarly, the 90-9-1 principle states that 90% of SNS actors observe and do not participate, 9% contribute sparingly, and 1% create the vast majority of new content. This 90%, 9%, and 1% are also known as Lurkers, Contributors, and Superusers. To date very little empirical research has been conducted to verify the 1% rule.

The purpose of this study was to examine if the 1% rule applied to moderated DHSNs designed to facilitate behavior change. Paid employees who were trained in social cognitive theory [23], motivational interviewing [24], the stages of change [25], and cognitive behavioral therapy (CBT) [26] actively moderated the four DHSNs in this study. Moderator roles focused on facilitating discussions, encouraging problem solving among members, administering compliance with privacy protection rules, protecting the community from spam, and ensuring that all discussions focused on adherence to behavior-change principles.

Furthermore, if DHSNs are efficacious, is it possible to create network effects to increase wellness on a population level? If size of the network matters, how are positive externalities created? More importantly, how do different actors interact and is it possible to create bandwagon effects?

Methods

Settings and Program Descriptions

To verify the 1% rule, this observational study analyzed descriptive data from four eHealth interventions that contain large social networks. The four Internet interventions are AlcoholHelpCenter (problem drinking) [27], DepressionCenter (depression) [28], PanicCenter (panic) [29], and StopSmokingCenter (smoking cessation) [30].

All four DHSNs are online, free to participants, do not offer advertising, do not promote any products, and are a part of Evolution Health Systems Inc's (EHS) social business model. EHS is a private, research-based organization that builds evidence-based digital programs designed to increase medication and treatment adherence. The four DHSNs analyzed in this study were originally built by EHS for research purposes.

During the study period, moderators consistently monitored each DHSN, reviewed all 578,349 DHSN posts, and checked for their accuracy and consistency. Posts that did not specifically address behavior change or comply with program rules were removed.

Full descriptions of each intervention appear elsewhere [31-34]. The oldest of the four DHSNs was nearly 11 years in operation at time of this study, and functionality of each DHSN has been enhanced over time. For explanatory purposes, Table 1 outlines the main features of each program.

Retrospective data were extracted from each program's structured query language (SQL) database. Descriptive statistics were analyzed in SPSS version 19 for Mac.

All data collection procedures adhered to international privacy guidelines [35-37] and were in accordance with the Helsinki Declaration of 1975, as revised in 2008 [38]. The study was consistent with the University Research Ethics Committee procedures at Henley Business School, University of Reading, and was exempt from full review.

Table 1. Program features and functionality.

	Problem drinking	Depression	Panic disorder	Smoking cessation
Moderated social network	✓	✓	✓	✓
Tailored behavior-change program	✓	✓	✓	✓
Brief intervention/screener	✓	✓	✓	✓
Blogs	✓	✓	✓	✓
Private messaging among members	–	✓	–	✓
Video testimonials	✓	✓	✓	✓
Public profile	✓	✓	✓	✓
Symptom diary/tracker	✓	✓	✓	✓
Gamification (techniques to increase usability leveraging desire for achievement, rewards, and competition)	–	✓	–	✓

Registrants and Study Duration

The four DHSNs had varying numbers of members and life spans (see Table 2). Periods of analysis ranged from 4.0 years (problem drinking) to 10.9 years (smoking cessation).

The dataset was purged of moderator accounts to ensure that all content originated only from registered members. Only registered members could actively contribute to discussions; however, registration was not required to read or review all existing or newly generated content.

Table 2. Subjects and study duration.

	Problem drinking	Depression	Panic disorder	Smoking cessation
Date of first post	July 25, 2008	April 5, 2003	January 7, 2002	September 17, 2001
Date of last post	August 7, 2012	August 5, 2012	August 7, 2012	August 7, 2012
Number of days	1474	3411	3866	3978
Years	4.0	9.3	10.6	10.9
Registrants, n	2597	5151	11,372	44,870

Results

Summary

Descriptive statistics revealed that less than 25% of actors in each DHSN authored one or more posts (see Table 3).

Post frequencies in each of the four DHSNs were divided into the top 1% (Superusers), the next 9% (Contributors), and the remaining 90% (Lurkers) of actors. Each DHSN revealed similar patterns, with Superusers generating 59.0%-75.0%, Contributors authoring 23.8%-37.4%, and Lurkers only creating 1.1%-7.8% of all posts (see Figure 1 and Table 4).

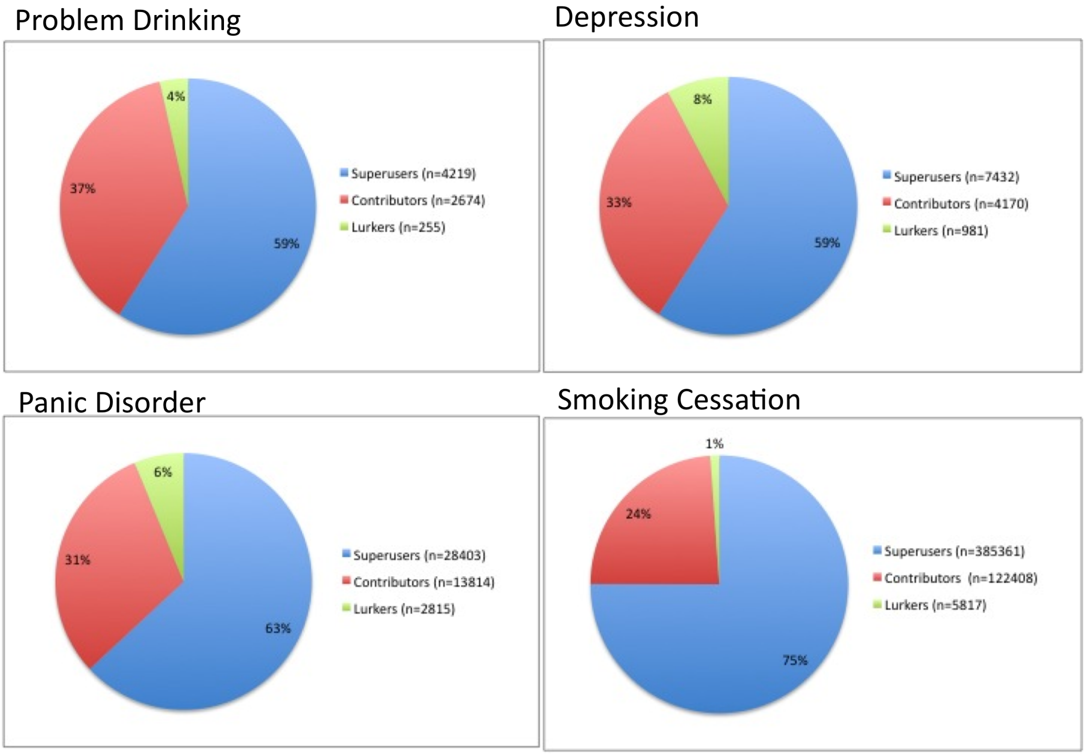
Table 3. Number and percentage of actors making one or more posts.

	Problem drinking	Depression	Panic disorder	Smoking cessation	Total	Mean	Weighted mean
Total actors	2597	5151	11,372	44,870	63,990	15,998	31,934
Actors who made at least one post, n (%)	449 (17.3)	1230 (23.9)	2767 (24.3)	7963 (17.7)	12409 (19.4)	3102 (19.4)	6193 (19.4)

Table 4. Analysis of the 1% rule.

	Problem drinking	Depression	Panic disorder	Smoking cessation	Total	Mean	Weighted mean
Total population (n)	2597	5151	11,372	44,870	63,990	15,998	40,875
Total social network posts	7148	12,583	45,032	513,586	578,349	144,587	369,434
1% of population (Superusers)	26	52	114	449	641	160	415
Total posts by Superusers	4219	7,432	28,403	385,361	425,415	106,354	276,034
Percentage of posts by Superusers, %	59.0	59.1	63.1	75.0	73.6	73.6	74.7
9% of population (Contributors)	234	464	1023	4038	5759	2880	3572
Total posts by Contributors	2674	4,170	13,814	122,408	143,066	35,767	88,732
Percentage of posts by Contributors, %	37.4	33.1	30.7	23.8	24.7	24.7	24.0
90% of population (Lurkers)	2337	4636	10,235	40,383	57,590	14,398	27,246
Total posts by Lurkers	255	981	2815	5817	9868	2467	4668
Percentage of posts by Lurkers, %	3.6	7.8	6.3	1.1	1.7	1.7	1.3

Figure 1. Network content according to the 1% rule.



The 1% (Superusers)

On average, the top 1% (n=160) of Superusers created 73.6% (n=106,354) of posts. On an individual program level, the top 1% varied in their overall contributions, but in all cases accounted for the majority of activity, with a weighted average of posts being 74.7% (n=276,034).

The Next 9% (Contributors)

The second highest group of contributors, or the next 9% of the population, accounted for an average of 24.7% (n=35,767), with a weighted average of 24.0% (n=88,732) of posts.

The Remaining 90% (Lurkers)

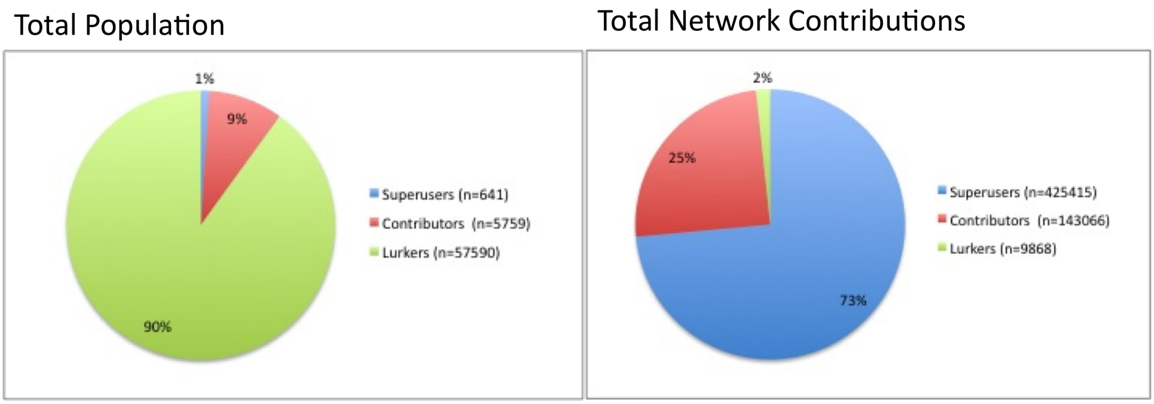
The remaining 90% of the population accounted for an average of 1.7% (n=2467) of posts, with a weighted average of 1.3% (4668) of posts.

Cumulative Participation

Cumulatively, Lurkers accounted for the vast majority of the population in the four DHSNs (n=57,590); however, this population created only 1.7% (n=9868) posts. Conversely, Superusers accounted for a small amount of actors (n=641) but created 73.6% (n=425,415) posts (see Figure 2).

Based on the overwhelming creation of content from a small number of Superusers and underwhelming amount of number of posts from a large number of lurkers, cyberculture’s 1% rule applies to the creation of positive network externalities in the four DHSNs analyzed in this study.

Figure 2. Cumulative DHSN population distribution and content creation according to the 1% rule.



Discussion

Principal Findings

Superusers accounted for a weighted average of 74.7% of content and generated the vast majority of posts within the four DHSNs. These findings match criteria of the 1% rule and may be comparable to the Pareto Principle.

Conversely, Lurkers generated limited or no network value. Although Lurkers may benefit from observing interactions between Superusers and Contributors, they do not generate network effects nor do they contribute to the network growth.

In regards to Superuser participation, motivations and posting patterns in the DHSNs have been previously examined. A 2008 analysis of the problem drinking DHSN found that common themes included introductions, greetings, general supportive statements, suggested strategies, success stories, and discussion of difficulties [9]. In addition, this study found that the amount of discussions varied over time and clustered around nodes consisting of one or more Superusers. A 2010 publication on the smoking cessation DHSN found that the majority of first posts were from recent quitters who were struggling with their quit attempts. Responses were rapid and from seasoned quitters, indicating that the social network may be particularly beneficial for peer support to help relapse prevention [11].

Content analysis has also been conducted on the four DHSNs. A 2009 academic presentation found that a high proportion of first posts in the panic disorder DHSN resembled “panic stories”, suggesting that the network may act as an expressive writing forum [39]. A 2010 academic presentation on the same community found that the support group was used more often by those reporting greater intensity of panic symptoms, absenteeism from work, and that Lurkers completed a greater number of the program’s CBT treatment sessions compared to Contributors and Superusers [40]. A recent University of Toronto PhD dissertation found that depression DHSN users generally sought informational support, various types of emotional support, coaching support, and social companionship [41]. Future research should focus on possible differences between post frequencies and content themes that may be prevalent in different indications, disease states, or actor types.

Based on the observations in this study, health care organizations should focus efforts on recruiting and retaining Superusers. Superusers may have a wide range of options to focus their participation, whether on health-related social networks or those of general interest. Moreover, they may exhibit different patterns of network behavior in different communities [42]. The motivations, needs, and participatory patterns of Lurkers and Contributors should also be examined. Future research should focus on the demographic and psychographic characteristics of these three actor-types.

It is also important to consider that the actions of some Superusers may result in negative network externalities. This type of behavior may result in negative network effects and decrease the size of the network. Conversely, Superusers may generate positive network effects in digital resources that are negatively oriented towards health, promoting illness, or disease [43].

An increasing number of health care organizations are making digital health care tools available to their patients, policyholders, or consumers, and many of them contain social networks. While some DHSNs flourish, many suffer from little or no traffic [44]. Strategies increasing Superuser and Contributor participation can increase the effectiveness of these programs.

A successful DHSN requires active managers who not only guide discussions but also facilitate growth [45]. The findings from this paper indicated that managers of DHSNs should identify Superusers early, encourage their participation, and target their recruitment though offline initiatives. Managers should not expend resources on promoting engagement with Lurkers.

Strengths and Limitations

A strength of this study is that the four DHSNs have never been promoted or advertised as they are not commercial entities. Participants in the four programs in this study could find the DHSNs only through extensive search efforts, links from other websites, or word-of-mouth. Profit-driven commercial entities focus considerable efforts and budgets on recruitment and promotion (free trials, banner advertising, celebrity endorsement, offline promotion, and other incentives) and most likely attract much larger populations with different motivations [46]. As a result of non-promotion, the four DHSNs in this study may have attracted only naturalistic, self-seeking health populations.

However, lack of advertising or promotion may also be a limitation. The naturalistic self-seeking population of actors within these networks may not be representative of populations that are typically reached from well-promoted programs. Many organizations or trials have promotional or recruitment budgets, thus casting a wider net and attracting a variety of health populations.

Especially in a climate of limited budgets and funding, the influence of promotion or non-promotion should encourage organizations with DHSNs to carefully consider the role of advertising and recruitment, and if those efforts should be strategically targeted.

Finally, only data from registered users were examined. Any visitor could browse the DHSNs without registering, but it is not possible to reliably examine this data nor combine it with the behavior of registered users.

Conclusions

The 1% rule was consistent across the four DHSNs. However, as individuals can lurk without registering, the 1% (Superusers) may represent an even smaller population. As social network sustainability requires fresh content and timely interactions, these results are important for organizations actively promoting and managing DHSNs.

Superusers generate the vast majority of traffic and create value, so their recruitment and retention is imperative for long-term success. Although Lurkers may benefit from observing interactions between Superusers and Contributors, they generate limited or no network value.

The results of this study indicate that DHSNs have the potential to be optimized to produce network effects, positive externalities, and bandwagon effects. Further research in the development, expansion, and management policies of DHSNs is required.

Acknowledgments

The author would like to thank Dr Douglass Hyatt, Professor of Business Economics at the Rotman School of Management at the University of Toronto for his input on network effects and comments on the study design.

Conflicts of Interest

Trevor van Mierlo is the CEO & Founder of Evolution Health Systems Inc, the owner of the sites AlcoholHelpCenter, DepressionCenter, PanicCenter, and StopSmokingCenter, as well as other eHealth and mHealth platforms.

References

1. Robinson TN, Walters PA. Health-net: an interactive computer network for campus health promotion. J Am Coll Health 1986 Jun;34(6):284-285. [doi: 10.1080/07448481.1986.9938952] [Medline: 3745693]
2. Schneider SJ. Trial of an on-line behavioral smoking cessation program. Computers in Human Behavior 1986 Jan;2(4):277-286. [doi: 10.1016/0747-5632(86)90008-7]
3. Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. Am J Public Health 2010 Jul;100(7):1282-1289 [FREE Full text] [doi: 10.2105/AJPH.2009.165449] [Medline: 20466971]
4. 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]
5. 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]

6. Welbourne JLL, Blanchard AL, Wadsworth MB. Motivations in virtual health communities and their relationship to community, connectedness and stress. Computers in Human Behavior 2013 Jan;29(1):129-139. [doi: 10.1016/j.chb.2012.07.024]
7. Newman MEJ. The Structure and Function of Complex Networks. SIAM Rev 2003 Jan;45(2):167-256. [doi: 10.1137/S003614450342480]
8. Schneider A, von Krogh G, Jäger P. “What’s coming next?” Epistemic curiosity and lurking behavior in online communities. Computers in Human Behavior 2013 Jan;29(1):293-303. [doi: 10.1016/j.chb.2012.09.008]
9. Cunningham JA, van Mierlo T, Fournier R. An online support group for problem drinkers: AlcoholHelpCenter.net. Patient Educ Couns 2008 Feb;70(2):193-198. [doi: 10.1016/j.pec.2007.10.003] [Medline: 18022340]
10. Jones R, Sharkey S, Smithson J, Ford T, Emmens T, Hewis E, et al. Using metrics to describe the participative stances of members within discussion forums. J Med Internet Res 2011;13(1):e3 [FREE Full text] [doi: 10.2196/jmir.1591] [Medline: 21239373]
11. 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]
12. Zhao K, Greer G, Qiu B, Mitra P, Portier K, Yen J. Working Paper. 2012. Finding influential users of an online health community: a new metric based on sentiment influence URL: http://arxiv.org/pdf/1211.6086v2.pdf [accessed 2014-01-20] [WebCite Cache ID 6MmVRkzEW]
13. Binks M, van Mierlo T. Utilization patterns and user characteristics of an ad libitum Internet weight loss program. J Med Internet Res 2010;12(1):e9 [FREE Full text] [doi: 10.2196/jmir.1347] [Medline: 20350926]
14. Grabher G, Stark D. Organizing Diversity: Evolutionary Theory, Network Analysis and Postsocialism. Regional Studies 1997 Jul;31(5):533-544. [doi: 10.1080/00343409750132315]
15. Uzzi B. The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect. American Sociological Review 1996 Aug;61(4):674-698 [FREE Full text] [doi: 10.2307/2096399]
16. Leibenstein H. Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand. The Quarterly Journal of Economics 1950 May;64(2):183-207. [doi: 10.2307/1882692]
17. Katz ML, Shapiro C. Network Externalities, Competition, and Compatibility. The American Economic Review 1985;75(3):424-440.
18. Lin KY, Lu HP. Why people use social networking sites: An empirical study integrating network externalities and motivation theory. Computers in Human Behavior 2011 May;27(3):1152-1161. [doi: 10.1016/j.chb.2010.12.009]
19. Sanders R. The Pareto Principle: its Use and Abuse. Journal of Product & Brand Management 1992;1(2):37-40. [doi: 10.1108/10610429210036762]
20. Nielsen J. Participation inequality: lurkers vs contributors in internet communities. URL: http://www.nngroup.com/articles/participation-inequality/ [accessed 2013-09-18] [WebCite Cache ID 6JjNtEa6S]
21. Arthur C. The Guardian. 2006 Jul 20. What is the 1% rule? URL: http://www.theguardian.com/technology/2006/jul/20/guardianweeklytechnologysection2 [accessed 2013-09-18] [WebCite Cache ID 6JjNf7fLJ]
22. Newman MEJ. Power laws, Pareto distributions and Zipf's law. Contemporary Physics 2005 Sep;46(5):323-351. [doi: 10.1080/00107510500052444]
23. McAlister L, Perry C, Parcel GS. How Individuals, Environments, and Health Behaviors Interact. In: Glanz K, Rimer BK, Viswanath K, editors. Health behavior and health education: theory, research, and practice. San Francisco: Jossey-Bass; 2008.
24. Miller WR, Rollnick S. Motivational Interviewing, Second Edition: Preparing People for Change. New York: The Guilford Press; 2002.
25. Freeman A, Dolan M. Revisiting prochaska and DiClemente's stages of change theory: An expansion and specification to aid in treatment planning and outcome evaluation. Cognitive and Behavioral Practice 2001 Jun;8(3):224-234. [doi: 10.1016/S1077-7229(01)80057-2]
26. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
27. Evolution Health Systems Inc. AlcoholHelpCenter.net. URL: http://www.alcoholhelpcenter.net/ [accessed 2014-01-20] [WebCite Cache ID 6MmZlwO8q]
28. Evolution Health Systems Inc. Depression Center.net. URL: http://www.depressioncenter.net/ [accessed 2014-01-20] [WebCite Cache ID 6MmZwBLeo]
29. Evolution Health Systems Inc. PanicCenter.net. URL: http://www.paniccenter.net/ [accessed 2014-01-20] [WebCite Cache ID 6Mma2bUXf]
30. Evolution Health Systems Inc. StopSmokingCenter.net. URL: http://www.stopsmokingcenter.net/ [accessed 2014-01-20] [WebCite Cache ID 6MmaX5xHy]
31. Cunningham JA. Comparison of two internet-based interventions for problem drinkers: randomized controlled trial. J Med Internet Res 2012;14(4):e107 [FREE Full text] [doi: 10.2196/jmir.2090] [Medline: 22954459]

32. Farvolden P, Cunningham JA, Selby P. Using E-Health Programs to Overcome Barriers to the Effective Treatment of Mental Health and Addiction Problems. Journal of Technology in Human Services 2009 Feb 03;27(1):5-22. [doi: 10.1080/15228830802458889]
33. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a Web-based self-help cognitive behavioral therapy program for panic disorder. J Med Internet Res 2005;7(1):e7 [FREE Full text] [doi: 10.2196/jmir.7.1.e7] [Medline: 15829479]
34. 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]
35. European Parliament and the Council of the European Union. Directive/58/EC of the European Parliament and of the Council of 12 July Concerning the Processing of Personal Data and the Protection of in the Electronic Communications Sector (Directive on Privacy and Electronic Communications). 2002. URL: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:HTML [accessed 2014-01-20] [WebCite Cache ID 6MmbBcYQM]
36. Office of the Privacy Commissiners of Canada. Personal Information Protection and Electronic Documents Act. 2006. URL: http://www.priv.gc.ca/legislation/02_06_01_e.cfm [accessed 2010-02-02] [WebCite Cache ID 5nEGDr6jp]
37. US Department of Health & Human Services. 2003. Summary of the HIPAA Privacy Rule URL: http://www.hhs.gov/ocr/hipaa/ [accessed 2008-05-08] [WebCite Cache ID 5Xfm36ydO]
38. Declaration of Helsinki: ethical principles for medical research involving human subjects. 2008. URL: http://www.wma.net/en/30publications/10policies/b3/index.html [accessed 2011-05-26] [WebCite Cache ID 5yxLF6EC6]
39. Hood H, Vorstenbach V, John B, Kloke S, Farvolden P, Anthony MM, et al. Tell Me About It! Content Analysis of Support Group Discussion on a Web-based Panic Disorder Treatment Program. 2009 Presented at: Annual meeting of the Association for Behavioral and Cognitive Therapies; November 19-22, 2009; New York, NY.
40. Vorstenbach V, Hood H, Farvolden P, Anthony MM, van Mierlo T. The Utility of an Online Support Group for Treatment Adhernece to a Web-Based CBT Program for Panic Disorder. 2010 Presented at: World Congress of Behavioral and Cognitive Therapies; June 2-5, 2010; Boston, MA.
41. Sugimoto S. Support Exchange on the Internet: A Content Analysis of an Online Support Group for People Living with Depression. In: Unpublished doctoral dissertation. Toronto: University of Toronto; 2013.
42. van Mierlo T, Fournier R, McGowan A, Selby P. All Superusers are not created equal: contributory patterns observed in four separate digital health social networks proming health behavior change. 2013 Presented at: Medicine 2.0: World Congress on Social Media, Mobile Apps, Internet / Web 2.0; September 23-24, 2013; London, UK.
43. Jade. Proana: Anorexia is a lifestyle, not a disease. 2013. URL: http://proanalifestyle.blogspot.ca/ [accessed 2014-01-20] [WebCite Cache ID 6MmcjzcOo]
44. Bender JL, Jimenez-Marroquin MC, Ferris LE, Katz J, Jadad AR. Online communities for breast cancer survivors: a review and analysis of their characteristics and levels of use. Support Care Cancer 2013 May;21(5):1253-1263. [doi: 10.1007/s00520-012-1655-9] [Medline: 23179491]
45. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]
46. Denos M. MarketingProfs. 2013. A Weight Watchers Case Study: How Smart Marketing Pays Off URL: http://www.marketingprofs.com/articles/2013/10422/a-weight-watchers-case-study-how-smart-marketing-pays-off [accessed 2014-01-20] [WebCite Cache ID 6MmdWWyRU]

Abbreviations

AHC: AlcoholHelpCenter.net
CBT: cognitive behavioral therapy
DC: DepressionCenter.net
DHSN: digital health social networks
PC: PanicCenter.net
SNS: social networking sites
SSC: StopSmokingCenter.net
SQL: structured query language

Edited by H Riper; submitted 18.09.13; peer-reviewed by D De Beurs, S McIntosh; comments to author 11.11.13; revised version received 21.12.13; accepted 10.01.14; published 04.02.14.

Please cite as:
van Mierlo T
The 1% Rule in Four Digital Health Social Networks: An Observational Study
J Med Internet Res 2014;16(2):e33
URL: http://www.jmir.org/2014/2/e33/
doi:10.2196/jmir.2966
PMID:24496109

Original Paper

Emotional Coping Differences Among Breast Cancer Patients From an Online Support Group: A Cross-Sectional Study

Anika Batenburg[1*], MSc; Enny Das[2*], PhD

[1]VU University Amsterdam, Department of Communication Science, Amsterdam, Netherlands

[2]Radboud University Nijmegen, Centre for Language Studies, Nijmegen, Netherlands

[*]all authors contributed equally

Corresponding Author:
Anika Batenburg, MSc
VU University Amsterdam
Department of Communication Science
De Boelelaan 1081
Amsterdam, 1081 HV
Netherlands
Phone: 31 20 5987 034
Fax: 31 20 598 3733
Email: a.e.batenburg@vu.nl

Abstract

Background: Due to mixed findings in research on the effect of online peer-to-peer support on psychological well-being, there is a need for studies explaining why and when online support communities are beneficial for cancer patients.

Objective: Previous studies have typically not taken into account individual coping differences, despite the fact that patients have different strategies to cope with cancer-related emotions. In the current study, it was predicted that the effects of online support group participation would partly depend on patients’ ability to cope with thoughts and emotions regarding the illness.

Methods: For this study, 184 Dutch breast cancer patients filled out a questionnaire assessing activity within a peer-led online support community, coping with emotions and thoughts regarding the illness (cognitive avoidance, emotional processing, and expression) and psychological well-being (depression, breast cancer-related concerns, and emotional well-being). Of these, 163 patients were visiting an online peer-led support community.

Results: Results showed interactions of the intensity of support group participation and coping style on psychological well-being. Specifically, we found an interaction of online activity and emotional expression on depression (beta=–.17, *P*=.030), a marginally significant interaction of online activity and emotional expression on emotional well-being (beta=.14, *P*=.089), and an interaction of online activity and cognitive avoidance on breast cancer–related concerns (beta=.15, *P*=.027). For patients who actively dealt with their emotions and thoughts, active online support group participation was positively related to psychological well-being. For patients high on avoidance of illness-related thoughts or low on emotional expression, active participation was negatively related to measures of well-being.

Conclusions: The current study revealed the role of individual differences in coping in online support group participation. Results suggest that breast cancer patients’ ability to cope with emotions and thoughts regarding the illness influence the relationship between online support group participation and psychological well-being.

(J Med Internet Res 2014;16(2):e28) doi:10.2196/jmir.2831

KEYWORDS

Internet; support groups; self-help groups; social support; online systems; emotions; coping style

Introduction

As in other countries, the number of newly diagnosed breast cancer patients in the Netherlands has increased in recent years (up 18% between 2000 and 2010; 13,357 new patients in 2010) and is expected to rise because of the extended life span of the population, the early detection of breast cancer, and the increasing effectiveness of treatment [1]. A greater number of patients are turning to the Internet for support and for information about health-related issues [2,3]. Research has shown that breast cancer patients are among the most active seekers of online social support compared to other patient groups [4]. Several studies have set out to reveal the effects of online peer support on psychological well-being, showing mixed findings [5-7]. For example, some studies revealed positive outcomes, such as decreased depression, reactions to pain, cancer-related trauma, and distress [8-10]. Other studies found no increase in well-being [11,12] or reported negative associations or effects [13,14].

These mixed findings may stem from the fact that most outcome studies assess online support interventions set up by professionals. These interventions often include other therapeutic elements besides peer support, which makes it hard to disentangle the effects of solely the online peer support [6]. Studies testing the effects of peer-led online support communities are underrepresented in the literature, even though these communities are common and easily accessible online. One reason might be that testing the effects of participating in peer-led communities is complicated due to its uncontrolled setting. Therefore, studies examining these peer-led groups are largely qualitative or descriptive in nature and suggest the presence of empowering processes, such as emotional and informational support, emotional expression, advice, recognition, understanding, and insight [15-19]. However, follow-up research is needed to assess the relationship between these therapeutic processes and health outcomes. Therefore, the first goal of our research was to provide a quantitative test of the relationship between participation and well-being in peer-led support groups.

Another reason for mixed findings in outcome studies is the individual differences between patients. Although descriptive studies suggest therapeutic processes at work within online support communities, these processes may not apply equally to all patients. Some patients only read stories from others, while others share experiences, thoughts, emotions, ask questions, or support others [20]. Although patients have the opportunity to disclose their feelings and read about the experiences and emotions of others online, not every patient approaches emotions equally. Previous descriptive studies show a large variety in content of online messages, from emotional experiences to illness to unrelated chitchat [21,22]. Furthermore, studies using a word counting program have shown that there are variations in individual writing styles and that these variations are related to health outcomes. For example, greater expression of negative emotions, cognitive processing, and lower expression of health-related concerns was positively associated with quality of life variables [11], and words suggestive of learning or understanding improved emotional well-being, functional well-being, health self-efficacy, and reduced negative mood [23]. Thus far, individual differences in how patients cope with thoughts and emotions have not been connected to potential effects of support group participation. A second goal of the present study therefore was to examine the relationship between peer-led support group participation, individual differences in coping behavior, and measures of well-being.

The intensity of support group use and how patients disclose thoughts and emotions depend to a large extent on individual coping styles. From offline studies, we know that some breast cancer patients approach the illness actively by processing and expressing their emotions [24], while others try to avoid all thoughts and emotions related to the illness [25]. The tendency to approach thoughts and emotions can be categorized on three levels of emotional engagement: (1) cognitive avoidance: the patient tries to avoid all thoughts and emotions related to the illness [26], (2) emotional processing: the attempt to acknowledge and understand feelings, but not necessarily the expression of emotions [24,27], and (3) emotional expression: individuals not only acknowledge their emotions but also allow themselves to express them [24]. Several studies have shown that the use of a cognitive avoidant coping style is negatively related to adjustment to cancer [28], health status [29], and is positively related to higher distress levels [25,30]. In contrast, actively dealing with thoughts and emotions regarding breast cancer showed a positive relationship with psychological and physical well-being, such as decreased depressive symptoms, distress, increased vigor, improved perceived health status, and fewer medical appointments [24,31-33]. Therefore, consistent with the coping literature, we expect that patients who score relatively high on avoidance coping will report a lower sense of well-being than patients who show less avoidant behavior. In contrast, patients who actively approach their emotions are expected to report a better sense of well-being than patients who approach emotions less actively.

Although one would expect that patients participating in an online support community are willing to confront illness-related information and therefore cope with their emotions and thoughts quite actively, a certain degree of online avoidance, processing, and expression might influence patients' well-being. Based on previous findings, it seems plausible to assume that active online support group participation may positively affect well-being in particular for patients who score low on cognitive avoidance and high on emotional processing and expression coping, because these patients are likely to benefit most from interactions with peers. Therefore, we suggest that coping styles are not only directly related to psychological well-being, but they also moderate the relationship between online support group participation and psychological well-being.

Due to the lack of studies on peer-led online support groups, we conducted a study on Dutch online communities set up by patients and former patients. Because previous mixed findings in research might partly have been caused by individual differences, this is the first attempt to include patients' coping style regarding illness-related thoughts and emotions. Although psychological research has extensively shown the relationship between coping styles and health outcomes [25,27-33], to our knowledge no study has assessed this relationship in the context of online peer support. We conducted our online study among Dutch breast cancer patients to measure the intensity of their online support group participation, coping style, cognitive avoidance, emotional processing, and emotional expression, as well as three measures of well-being generally associated with breast cancer diagnosis: emotional well-being, depression, and breast cancer-related concerns [34-36]. We also included additional factors in our analyses often associated with the psychological well-being of breast cancer patients, such as social support from family and friends [37,38], illness stage, and professional psychological help received. We hypothesized that active online support group participation would be positively related to emotional well-being and negatively related to depression and breast cancer-related concerns in particular for patients who scored low on cognitive avoidance and high on emotional processing and expression.

Methods

Participants and Procedure

We used Google to identify all online support communities for breast cancer patients in the Netherlands. Criteria for inclusion were that (1) the website was in the Dutch language, (2) the support group (sometimes part of a more extensive website) was designed as a peer-led message board available 24/7, and (3) the discussion board was still active (the last month's new messages had been posted). With approval of the website owners, a request to participate in an online survey about breast cancer patients' Internet use was posted on seven support websites (June 2011). This survey was part of a more extensive research project on online peer support among Dutch breast cancer patients. The research was carried out in accordance with the American Psychological Association's ethics guidelines [39] and complies with European Union legislation [40] and Dutch legislation [41] on data protection.

The introduction page (ie, the first page of the survey) included the length and purpose of the survey, ensured anonymity, and contact information of the investigator (in case participants had any questions). The first page of the survey was viewed 311 times, and 184 Dutch breast cancer patients filled out the questionnaire (182 females and 2 males). Response rates are unknown because we had no access to page views of the participating websites. The online survey tool tracked IP (Internet protocol) addresses to prevent users from re-taking the survey. Responses to questions were obligatory, but participants were provided with an “I don't know” or “not applicable” option.

Since males were underrepresented, we decided to exclude them from data analysis. Another 7 participants were excluded, due to extreme responses on one of the dependent variables (SD>3). Therefore, 175 participants were included in data analysis. Table 1 shows the demographics and patient characteristics of the study sample. Table 2 shows the average use of the peer support message board.

Table 1. Demographics and health characteristics (N=175).

Characteristics		n	%
Age			
Mean (SD)		48.09 (9.04)	
Minimum		23	
Maximum		71	
Education [a]			
Elementary school		5	2.9
Secondary education			
	Low	29	16.6
	Middle	15	8.6
	High	2	1.1
Tertiary education			
	Low[b]	10	5.7
	Middle	47	26.9
	High	55	31.4
	Scientific degree	12	6.9
Working status			
Not working		90	51.4
Working		85	48.6
Illness stage			
No cancer cells at the moment		27	15.4
Stage I (tumor smaller than 2 cm, no metastases to the lymph nodes)		41	23.4
Stage II (metastases to the lymph nodes in the armpit, or a tumor larger than 2 cm with no metastases)		42	24.0
Stage III (metastases to multiple lymph nodes or other lymph nodes)		25	14.3
Stage IV (metastases to other body parts)		16	9.1
Unknown		24	13.7
Psychological help during period of illness			
Yes		69	39.4
No		106	60.6

[a] Levels within the Dutch education system: education is divided over three schools for different age groups, which are divided in streams for different educational levels.

[b] Dutch educational structure “LBO/LTS” existed until 1992.

Table 2. Use of the online support community (N=175).

Frequency of use	n	%
Not visiting an online BC support message board	12	6.9
Frequency of visits (n=163)		
<1 per month	19	11.7
Approximately once a month	12	7.4
Multiple times per month	12	7.4
Approximately once a week	17	10.4
Multiple times per week	22	13.5
Approximately once a day	44	27.0
Multiple times per day	37	22.7
Frequency of posts the last 4 weeks (n=163)		
None	61	37.4
≤1 per week	39	23.9
Multiple posts per week, but not every day	45	27.6
Every day one post or more	18	11.0
Forum contribution (n=163)		
I only read posts from others	36	22.1
I reacted on post(s) of someone else	28	17.1
I started a new topic or asked a question	18	11.0
I both started a new topic or asked a question AND I reacted on post(s) of another	81	49.7
Average length of visits (n=163)		
<10 minutes	61	37.4
10-30 minutes	71	43.6
30 minutes to 1 hour	22	13.5
>1 hour	9	5.5

Measurements

Online Support Group Participation

The intensity of online support group participation was assessed by four different questions regarding frequency of visits, average length of visits, contribution, and frequency of posts in the last 4 weeks (cf [42]). Frequency of visits was assessed on a 7-point scale; the other items were assessed on a 4-point scale (Table 2). To merge these different scales into one index, all items were transformed into Z scores (Cronbach alpha=.77).

Emotional Coping

The Dutch mini-MAC [24] was used to assess *cognitive avoidance* (4 items, eg, "I try not to think about my illness"). Participants rated on a 4-point scale if the statements applied to them. Ratings were summed and averaged across items. Higher scores indicated that the coping style of cognitive avoidance applied to them. The scale was internally consistent (Cronbach alpha=.82).

The Emotional Approach Coping scale [22] was used to measure *emotional processing* and *emotional expression*. Four items measured emotional expression (eg, "I take the time to express my emotions"; Cronbach alpha=.86). Another four items measured emotional processing (eg, "I realize that my feelings are justified and important"; Cronbach alpha=.69). Participants rated on a 4-point scale if the statements applied to them. Ratings were summed and averaged across items. The variables indicating the intensity of support group participation and coping styles were unrelated.

Psychological Well-Being

Three different scales measured psychological well-being. First we measured depression using the Center for Epidemiological Studies Depression Scale (CES-D10) [43]. The scale consisted of 10 items (eg, "I felt that everything I did took me quite a lot of effort"). Participants rated on a 4-point scale if the statements applied to them the last week, from "less than one day" to "5 to 7 days". Ratings were summed and averaged across items. Higher scores indicated more depression-related thoughts. The scale was internally consistent: Cronbach alpha=.72. Additionally, we measured breast cancer-related concerns [44] with 28 items (eg, "Are you concerned that your friends will avoid you?"; Cronbach alpha=.89). Participants answered these questions on a 5-point scale, ranging from "Not at all" to "Totally". Higher scores indicated more concerns regarding the illness. Emotional well-being was measured according to 6 items from the Functional Assessment of Chronic Illness Therapy questionnaire (FACIT-B) [45] (Cronbach alpha=.81). An item example is “I feel sad”. Respondents rated on a 5-point scale if the statements applied to them, ranging from “Not at all” to “Totally”. Higher scores indicated a better sense of emotional well-being.

Control Variables

Last, we measured participants’ age, education level, current working status (ie, if they were currently working), illness stage (the standard four phases in breast cancer [46]), psychological help (ie, if they received psychological help from a professional), and offline social support (based on the six “Social well-being” items from the FACIT-B [45]). Items referring to support from friends were adjusted into items that clearly referred to their offline friends, not online peers. Respondents rated on a 5-point scale if the statements applied to them, ranging from “Not at all” to “Totally” (Cronbach alpha=.94).

Results

Correlations

Tables 3 and 4 show the correlations between all variables. Cognitive avoidance coping was associated with all three psychological well-being variables: positively with depression and breast cancer–related concerns and negatively with emotional well-being. Emotional processing was not related to psychological well-being. Emotional expression was negatively related to depression. No direct associations between support group participation and one of the psychological well-being variables were found.

We tested the hypothesized relationship between online support group participation, coping styles, and psychological well-being with regression analyses. All independent variables were standardized into Z scores to meet the requirements to perform regression analyses and to compute interaction terms. In addition, every model included the covariates that were significantly correlated to the dependent variable (ie, depression, emotional well-being, breast cancer-related concerns). We examined interactions for participants with relative low scores (1 SD below the mean of the standardized score) and for participants with relative high scores (1 SD above the mean of the standardized score) on the continuous indices measuring intensity of online participation and coping styles (see [47] for this regression analysis).

Table 3. Means, standard deviations, and intercorrelations of independent variables, illness stage, age, and dependent variables.

Variables		n	M	SD	1	2	3	4	5	6
1	Support group participation[a]	163	–0.01	0.78	–					
2	Cognitive avoidance	175	1.95	0.60	.03	–				
3	Emotional processing	174	2.86	0.50	.01	–.10	–			
4	Emotional expression	174	2.79	0.54	.07	–.16[d]	.57[c]	–		
5	Illness stage	151	2.75	1.23	.23[c]	.06	–.08	–.09	–	
6	Age	175	48.09	9.04	–.24[c]	–.02	–.03	.02	–.01	–
7	Education	175	6.40	1.93	.01	–.13	.09	.03	.11	–.26[c]
8	Working status[b]	175	0.49	0.81	–0.12	.01	–.08	–.14	–.13	–.20[c]
9	Offline social support	175	3.78	0.60	.03	–.18	–.02	.18[d]	–.03	–.05
10	Psychological help[b]	175	0.39	0.49	–.06	.11	.15	–.04	.00	–.22[c]
11	Depression	175	1.83	0.45	.01	.36[c]	.05	–.15[d]	.04	–.06
12	Breast cancer-related concerns	175	2.64	0.55	.08	.35[c]	.03	–.02	.07	–.18[d]
13	Emotional well-being	175	3.53	0.77	–.02	–.47[c]	–.06	.06	–.20[d]	.04

[a] Standardized into Z scores.

[b] Coded 0=no, 1=yes.

[c] Correlations significant at the .01 level.

[d] Correlations significant at the .05 level.

Table 4. Means, standard deviations, and intercorrelations of education level, working status, psychological, and offline support, and dependent variables.

Variables		n	M	SD	7	8	9	10	11	12
1	Support group participation[a]	163	–0.01	0.78						
2	Cognitive avoidance	175	1.95	0.60						
3	Emotional processing	174	2.86	0.50						
4	Emotional expression	174	2.79	0.54						
5	Illness stage	151	2.75	1.23						
6	Age	175	48.09	9.04						
7	Education	175	6.40	1.93	–					
8	Working status[b]	175	0.49	0.81	.27[c]	–				
9	Offline social support	175	3.78	0.60	.15[d]	.09	–			
10	Psychological help[b]	175	0.39	0.49	.13	.13	–.18[d]	–		
11	Depression	175	1.83	0.45	–.13	–.17[d]	–.36[c]	.24[c]	–	
12	Breast cancer-related concerns	175	2.64	0.55	–.09	–.19[d]	–.44[c]	.18[d]	.43[c]	–
13	Emotional well-being	175	3.53	0.77	.04	.20[c]	.32[c]	–.18[d]	–.56[c]	–.49[c]

[a] Standardized into Z scores.

[b] Coded 0=no, 1=yes.

[c] Correlations significant at the .01 level.

[d] Correlations significant at the .05 level.

Depression

Regression results (Table 5) indicated a main effect of working status, psychological help, offline social support, cognitive avoidance, and a marginally significant effect of emotional expression on depression. No main effects of the intensity of support group participation and emotional processing were found. The higher that patients scored on cognitive avoidance, the more depressive feelings they reported. In contrast, emotional expression was negatively related to depression. Furthermore, an interaction effect of emotional expression and intensity of support group participation was found.

For patients expressing their emotions, intensity of online support group participation was associated with less depressive feelings. In contrast, for patients who scored low on emotional expression, online activity was associated with more depressive feelings (Figure 1). No other interaction effects on depression were observed.

Table 5. Hierarchical regression results for the effects of support group participation and coping style on depression, emotional well-being, and breast cancer–related concerns.

	Depression (n=162)			Emotional well-being (n=141)			Breast cancer concerns (n=162)		
Variable	b (SE)	ß	*P*	b (SE)	ß	*P*	b (SE)	ß	*P*
Step 1: Covariates									
Working status[a]	–0.13 (0.06)	–.14	.047	0.20 (0.11)	.13	.072	–0.22 (0.07)	–.20	.003
Psychological help[a]	0.14 (0.07)	.15	.039	–0.14 (.12)	–.88	.245	0.10 (0.08)	.09	.205
Offline social support	–0.10 (0.03)	–.23	.002	0.11 (0.06)	.15	.045	–0.22 (0.04)	–.42	<.001
Illness stage	–	–	–	–0.13 (0.06)	–.17	.023	–	–	–
Age	–	–	–	–	–	–	–0.14 (0.04)	–.25	<.001
Step 2: Main effects									
(A) Support group participation	0.01 (0.03)	.02	.748	–0.02 (0.06)	–.02	.778	0.03 (0.04)	.05	.459
(B) Cognitive avoidance	0.13 (0.03)	.28	.000	–0.37 (0.06)	–.47	<.001	0.14 (0.04)	.27	<.001
(C) Emotional processing	0.05 (0.04)	.11	.220	–0.07 (0.07)	–.09	.312	–0.07 (0.05)	–.12	.159
(D) Emotional expression	–0.07 (0.04)	–.16	.067	0.04 (0.07)	.05	.603	0.05 (0.04)	.09	.290
Step 3: Interaction effects									
A x B	–0.03 (0.03)	–.08	.273	–0.02 (0.05)	–.03	.709	0.07 (0.03)	.15	.027
A x C	–0.01 (0.04)	–.02	.856	0.09 (0.07)	.11	.167	–0.03 (0.04)	–.05	.529
A x D	–0.08 (0.04)	–.17	.030	0.10 (0.06)	.14	.089[b]	–0.02 (0.04)	–.03	.671
R^2	.31			.39			.41		
ANOVA	$F_{10,151}$=6741		<.001	$F_{11,129}$=7500		<.001	$F_{11,150}$=9309		<.001
Adjusted R^2	.26			.34			.36		
Cohen’s f^2	0.45			0.64			0.69		

[a] Coded 0=no, 1=yes.

[b] This marginally significant effect may be due to a smaller sample (n=141) caused by incomplete scores regarding the covariate *illness stage*; the interaction is significant without the inclusion of *illness stage*, b(SE)=0.17(.07), *ß*=.17, *P*=.02, n=162.

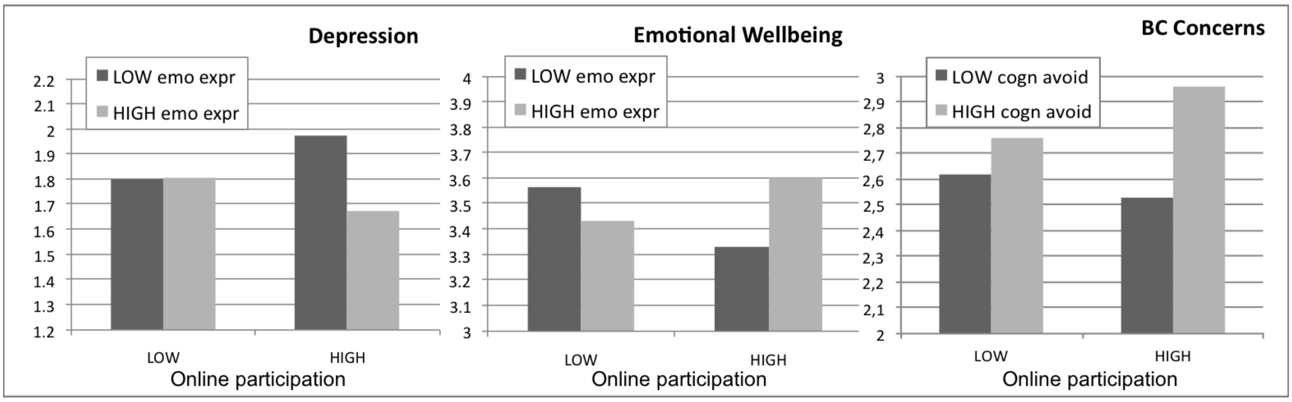


Figure 1. Interactions of the intensity of online support group participation and coping style on depression, emotional well-being, and breast cancer-related concerns (psychological well-being: Y-axis shows the absolute score; intensity of online participation and coping style: Low=1 standard deviation below the mean; High=1 standard deviation above the mean).

Emotional Well-Being

Offline social support, illness stage, and cognitive avoidance showed a main effect on emotional well-being. No main effects of intensity of online support group participation and the other coping styles were found. Again, the more patients avoided thoughts, the worse their sense of well-being. Furthermore, a marginally significant interaction between emotional expression and support group participation was found (Table 4).

Results tentatively suggest that for patients low on emotional expression, being more active in an online support group was negatively associated with emotional well-being. For patients who expressed their emotions, the intensity of online support group participation was positively related to emotional well-being (Figure 1). No other interaction effects were observed on emotional well-being.

Breast Cancer–Related Concerns

Results showed a main effect of working status, offline social support, age, and cognitive avoidance. No main effects of intensity of online participation or the other coping styles were found. Additionally, an interaction between cognitive avoidance and support group participation on breast cancer–related concerns was found (Table 4). Patients who were rather active online and tried to avoid thoughts about their illness had more breast cancer–related concerns than active patients who scored low on cognitive avoidance (Figure 1). No other interaction effects were found.

Discussion

Principal Findings

The present cross-sectional study tested the interaction of online support group participation and coping styles concerning illness-related thoughts and emotions on psychological well-being. The expected interaction was found on several occasions. Specifically, results suggest that patients coping with their illness by expressing their emotions may benefit more from online support group participation than patients who approach or acknowledge their emotions less. For patients who approached their emotions, active participation was positively related to emotional well-being and negatively related to depression. For patients who scored low on emotional expression, however, active participation was related to higher levels of depression and lower levels of emotional well-being. Finally, for patients who were more avoidant, the intensity of online support group participation was related to higher levels of breast cancer–related concerns, compared to patients who were less avoidant. These findings support the assumption that the relationship between online support group participation and well-being might be influenced by individual differences in coping styles. The current results suggest that active online support group participation may be more beneficial for individuals expressing their thoughts and emotions.

A potential explanation for the negative relation between the intensity of online participation and breast cancer-related concerns among patients with a more avoidant coping style is that these patients may be less able to cope with the negative content on online forums; they may be overwhelmed by the sad and frightening stories from patients in the same condition. In some online support group interview studies, patients mentioned having difficulties being confronted with negative sides of the disease [21,48], and some even withdrew to avoid painful details about cancer [49]. Previous research also showed that breast cancer patients use optimistic stories from peers as a source for inspiration [50]. Patients with an avoidant coping style may search for such positive stories but at the same time encounter negative stories they cannot cope with. Alternatively, patients with more concerns might be more active support seekers, but in turn also become more avoidant to be able to deal with their own extreme emotions and potentially distressing stories from online peers. Sometimes avoiding thoughts and emotions might be beneficial in order to prevent becoming overwhelmed with negative information from others. Since this study has a cross-sectional design, we cannot draw conclusions on the direction of the relationships we found. Therefore, there is a need for longitudinal studies on peer-led support groups, including patients' coping styles. In addition, future research should focus on the content to reveal more insight into support group participation of patients with different coping styles, especially avoidant patients.

The present findings further previous research by showing that individual coping differences among online support seekers are likely to influence the relationship between online support group participation and psychological well-being. The current study might explain null findings in previous studies [11,12]. We found no direct relation between the intensity of participation and well-being, but we did find interactions of coping style and online participation on well-being. Support group participation may intensify certain positive processes that are already present in patients, such as the expression of emotions, but may also influence patients negatively when adaptive coping styles are less present. It is important to take these personal factors into account when we investigate the effectiveness of online support groups. Recently, studies have started looking into individual differences that may influence the effects of online support group participation. For example, a study showed different effects on emotional well-being depending on patients' level of health self-efficacy [51].

Our findings also showed that other factors, such as illness stage and offline social support, were sometimes more related to psychological well-being than coping style or the interaction of support group participation and coping style. This underscores the importance of including “offline” factors that affect patients' well-being when examining online support group effectiveness. Illness characteristics are often considered, but other factors outside the online support group that may influence well-being, such as support from friends and relatives or professional psychological help are often left out. Researchers should be careful not to exaggerate the effects of online peer support and distinguish effects directly caused by online participation, effects caused by other (offline) factors, and factors that could be strengthened by online support group participation (such as certain coping strategies). Future studies should also further examine the strength of these effects, as the effect sizes observed in the present research were rather modest.

Limitations and Future Research

A limitation of this research is the cross-sectional design, which warrants caution in interpreting the direction of causality. A longitudinal study is needed to test causal relationships. For example, it is possible that not being able to express one's emotions negatively affects well-being and that this decreased well-being, in turn, prompts patients to become more active online in order to find support. From a modeling or skills perspective [52], active participation in online support communities may help patients with more repressive coping styles to learn over time how to positively approach their illness and express their emotions. The latter effects are more likely to surface in longitudinal studies that track patients over an extended period of time. Future studies should therefore test the causal directions of the presently observed relationship and also examine whether patients becoming more active online can be beneficial for patients with repressive coping styles in the longer run.

Furthermore, no interactions of emotional processing and online participation on psychological well-being were found. This can be explained by the suggestion of Stanton and colleagues [30] that the effect of emotional processing depends on its contribution to emotional expression. When thoughts and feelings are not expressed, they may become ruminative, which may negatively affect well-being. Considering that the effectiveness of approaching emotions depended on the expression of emotions—and not on processing emotions—the scale measuring emotional expression may be a more important determinant of well-being. Future research should look further into the interrelations between different, yet related, coping styles.

Finally, knowledge in this field may be extended by studying differences between forum users and non-users. Questions of interest include which patients decide to participate in these online support groups and whether non-users should at any time be encouraged to participate.

Conclusion

The current findings tentatively suggest that the effectiveness of online peer support is partly influenced by individual differences in coping style regarding thoughts and emotions. Patients who actively coped with their emotions and thoughts and participated actively within an online support community reported a better sense of psychological well-being than online active patients who coped less actively with their emotions. Although the current findings are cross-sectional, one plausible interpretation might be that patients who actively approach their emotions may benefit most from online support group participation, because online writing may reinforce the effectiveness of active coping styles. Since we found no direct relation between the intensity of online participation and psychological well-being, but several interactions between online activity and coping styles on well-being, patients' initial coping abilities should be taken into account when examining the effectiveness of online peer support in future research.

Acknowledgments

We would like to thank the website owners and patients for their support and participation.

Conflicts of Interest

None declared.

References

1. IKNL: Integraal kankercentrum Nederland. Cijfers over kanker - Prevalentie. 2013. URL: http://www.cijfersoverkanker.nl/selecties/Dataset_2/img52849a55363ad [accessed 2013-11-14] [WebCite Cache ID 6L7WkDhw2]
2. Nagler RH, Gray SW, Romantan A, Kelly BJ, DeMichele A, Armstrong K, et al. Differences in information seeking among breast, prostate, and colorectal cancer patients: results from a population-based survey. Patient Educ Couns 2010 Dec;81 Suppl:S54-S62 [FREE Full text] [doi: 10.1016/j.pec.2010.09.010] [Medline: 20934297]
3. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
4. 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]
5. Hong Y, Peña-Purcell NC, Ory MG. Outcomes of online support and resources for cancer survivors: a systematic literature review. Patient Educ Couns 2012 Mar;86(3):288-296. [doi: 10.1016/j.pec.2011.06.014] [Medline: 21798685]
6. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
7. Klemm P, Bunnell D, Cullen M, Soneji R, Gibbons P, Holecek A. Online cancer support groups: a review of the research literature. Comput Inform Nurs 2003;21(3):136-142. [Medline: 12792194]
8. Lieberman MA, Golant M, Giese-Davis J, Winzlenberg A, Benjamin H, Humphreys K, et al. Electronic support groups for breast carcinoma: a clinical trial of effectiveness. Cancer 2003 Feb 15;97(4):920-925 [FREE Full text] [doi: 10.1002/cncr.11145] [Medline: 12569591]
9. 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]
10. Han JY, Shah DV, Kim E, Namkoong K, Lee SY, Moon TJ, et al. Empathic exchanges in online cancer support groups: distinguishing message expression and reception effects. Health Commun 2011 Mar;26(2):185-197 [FREE Full text] [doi: 10.1080/10410236.2010.544283] [Medline: 21318917]

11. 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]
12. 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]
13. Winefield HR, Coventry BJ, Pradhan M, Harvey E, Lambert V. A comparison of women with breast cancer who do and do not seek support from the Internet. Aust J Psychol 2003;55(1):30-34. [doi: 10.1080/0004953041233131284]
14. Salzer MS, Palmer SC, Kaplan K, Brusilovskiy E, Ten Have T, Hampshire M, et al. A randomized, controlled study of Internet peer-to-peer interactions among women newly diagnosed with breast cancer. Psychooncology 2010 Apr;19(4):441-446. [doi: 10.1002/pon.1586] [Medline: 19484712]
15. Fernsler JI, Manchester LJ. Evaluation of a computer-based cancer support network. Cancer Pract 1997;5(1):46-51. [Medline: 9128496]
16. Rodgers S, Chen Q. Internet community group participation: Psychosocial benefits for women with breast cancer. Journal of Computer-Mediated Communication 2005;10(4):5. [doi: 10.1111/j.1083-6101.2005.tb00268.x]
17. 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]
18. Setoyama Y, Nakayama K, Yamazaki Y. Peer support from online community on the internet among patients with breast cancer in Japan. Stud Health Technol Inform 2009;146:886. [Medline: 19593035]
19. Mo PK, Coulson NS. Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online HIV/AIDS support group. Cyberpsychol Behav 2008 Jun;11(3):371-374. [doi: 10.1089/cpb.2007.0118] [Medline: 18537512]
20. 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 [FREE Full text] [doi: 10.2196/jmir.1696] [Medline: 22204869]
21. van Uden-Kraan C, Drossaert C, Taal E, Lebrun C, Drossaers-Bakker K, Smit W, et al. Coping with somatic illnesses in online support groups: Do the feared disadvantages actually occur? Computers in Human Behavior 2008 Mar;24(2):309-324. [doi: 10.1016/j.chb.2007.01.014]
22. 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]
23. Shaw BR, Hawkins R, McTavish F, Pingree S, Gustafson DH. Effects of insightful disclosure within computer mediated support groups on women with breast cancer. Health Commun 2006;19(2):133-142. [doi: 10.1207/s15327027hc1902_5] [Medline: 16548704]
24. Stanton AL, Kirk SB, Cameron CL, Danoff-Burg S. Coping through emotional approach: scale construction and validation. J Pers Soc Psychol 2000 Jun;78(6):1150-1169. [Medline: 10870915]
25. Carver CS, Pozo C, Harris SD, Noriega V, Scheier MF, Robinson DS, et al. How coping mediates the effect of optimism on distress: a study of women with early stage breast cancer. J Pers Soc Psychol 1993 Aug;65(2):375-390. [Medline: 8366426]
26. Watson M, Law MG, Santos MD, Greer S, Baruch J, Bliss J. The Mini-MAC. Journal of Psychosocial Oncology 1994 Oct 31;12(3):33-46. [doi: 10.1300/J077V12N03_03]
27. Strongman KT. Emotional creativity. In: International review of studies on emotion. Chichester: Wiley; 1991.
28. Friedman L, Nelson D, Baer P, Lane M, Smith F. Adjustment to Breast Cancer. Journal of Psychosocial Oncology 1991 Mar 21;8(4):27-40. [doi: 10.1300/J077v08n04_02]
29. Epping-Jordan JE, Compas BE, Howell DC. Predictors of cancer progression in young adult men and women: avoidance, intrusive thoughts, and psychological symptoms. Health Psychol 1994 Nov;13(6):539-547. [Medline: 7889909]
30. Stanton AL, Snider PR. Coping with a breast cancer diagnosis: a prospective study. Health Psychol 1993 Jan;12(1):16-23. [Medline: 8462494]
31. Classen C, Koopman C, Angell K, Spiegel D. Coping styles associated with psychological adjustment to advanced breast cancer. Health Psychol 1996 Nov;15(6):434-437. [Medline: 8973923]
32. Stanton AL, Danoff-Burg S, Cameron CL, Bishop M, Collins CA, Kirk SB, et al. Emotionally expressive coping predicts psychological and physical adjustment to breast cancer. J Consult Clin Psychol 2000 Oct;68(5):875-882. [Medline: 11068973]
33. Stanton AL, Danoff-Burg S, Huggins ME. The first year after breast cancer diagnosis: hope and coping strategies as predictors of adjustment. Psychooncology 2002;11(2):93-102. [Medline: 11921325]
34. Spencer SM, Lehman JM, Wynings C, Arena P, Carver CS, Antoni MH, et al. Concerns about breast cancer and relations to psychosocial well-being in a multiethnic sample of early-stage patients. Health Psychol 1999 Mar;18(2):159-168. [Medline: 10194051]
35. Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. J Clin Oncol 1997 Mar;15(3):974-986. [Medline: 9060536]

36. Hann D, Winter K, Jacobsen P. Measurement of depressive symptoms in cancer patients: evaluation of the Center for Epidemiological Studies Depression Scale (CES-D). J Psychosom Res 1999 May;46(5):437-443. [Medline: 10404478]
37. Dukes Holland K, Holahan CK. The relation of social support and coping to positive adaptation to breast cancer. Psychology & Health 2003 Jan;18(1):15-29. [doi: 10.1080/0887044031000080656]
38. Talley A, Molix L, Schlegel RJ, Bettencourt A. The influence of breast cancer survivors' perceived partner social support and need satisfaction on depressive symptoms: a longitudinal analysis. Psychol Health 2010 Apr;25(4):433-449. [doi: 10.1080/08870440802582682] [Medline: 20397295]
39. APA Ethics Committee. 2010 Amendments to the 2002 "Ethical principles of psychologists and code of conduct". Am Psychol 2010;65(5):493. [doi: 10.1037/a0020168] [Medline: 20642307]
40. EU legislation on data protection. 2013 Oct 09. URL: http://ec.europa.eu/justice/data-protection/ [accessed 2013-10-09] [WebCite Cache ID 6KEjYUQEE]
41. Dutch legislation on data protection. 2013 Oct 09. URL: http://wetten.overheid.nl/BWBR0011468/geldigheidsdatum_11-07-2013 [accessed 2013-10-09] [WebCite Cache ID 6KEjjpGbK]
42. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
43. Irwin M, Artin KH, Oxman MN. Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). Arch Intern Med 1999;159(15):1701-1704. [Medline: 10448771]
44. Spencer SM, Lehman JM, Wynings C, Arena P, Carver CS, Antoni MH, et al. Concerns about breast cancer and relations to psychosocial well-being in a multiethnic sample of early-stage patients. Health Psychol 1999 Mar;18(2):159-168. [Medline: 10194051]
45. Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. J Clin Oncol 1997 Mar;15(3):974-986. [Medline: 9060536]
46. Oldenburg H, Vrancken Peeters M, Van Bohemen J. Het borstkanker boek. Amsterdam: Thoeris; 2007.
47. Aiken LS, West SG. Multiple regression: testing and interpreting interactions. Newbury Park, CA: Sage Publications; 1991.
48. Barak A, Boniel-Nissim M, Suler J. Fostering empowerment in online support groups. Computers in Human Behavior 2008 Sep;24(5):1867-1883. [doi: 10.1016/j.chb.2008.02.004]
49. Sandaunet AG. The challenge of fitting in: non-participation and withdrawal from an online self-help group for breast cancer patients. Sociol Health Illn 2008 Jan;30(1):131-144. [doi: 10.1111/j.1467-9566.2007.01041.x] [Medline: 18254837]
50. Helgeson VS, Taylor SE. Social comparisons and adjustment among cardiac patients. Journal of Applied Social Psychology 1993;23(15):1171-1195. [doi: 10.1111/j.1559-1816.1993.tb01027.x]
51. Namkoong K, Shah DV, Han JY, Kim SC, Yoo W, Fan D, et al. Expression and reception of treatment information in breast cancer support groups: how health self-efficacy moderates effects on emotional well-being. Patient Educ Couns 2010 Dec;81 Suppl:S41-S47 [FREE Full text] [doi: 10.1016/j.pec.2010.09.009] [Medline: 21044825]
52. Bandura A. Self-efficacy. In: The Corsini Encyclopedia of Psychology (4th Ed). New Jersey: John Wiley & Sons; 2010.

Abbreviations

CES-D: Center for Epidemiological Studies—Depression Scale
FACIT: Functional Assessment of Chronic Illness Therapy
SE: standard error

Edited by G Eysenbach; submitted 17.07.13; peer-reviewed by N van Uden-Kraan, C Barker; comments to author 29.08.13; revised version received 10.10.13; accepted 15.11.13; published 05.02.14.

Please cite as:
Batenburg A, Das E
Emotional Coping Differences Among Breast Cancer Patients From an Online Support Group: A Cross-Sectional Study
J Med Internet Res 2014;16(2):e28
URL: http://www.jmir.org/2014/2/e28/
doi:10.2196/jmir.2831
PMID:24499687

Original Paper

The Effectiveness of an Internet Support Forum for Carers of People With Dementia: A Pre-Post Cohort Study

Vicky McKechnie[1], DClinPsy; Chris Barker[1], PhD; Josh Stott[1], DClinPsy

[1] Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom

Corresponding Author:
Chris Barker, PhD
Department of Clinical, Educational and Health Psychology
University College London
Gower St
London, WC1E 6BT
United Kingdom
Phone: 44 20 7679 5962
Fax: 44 29 7916 1989
Email: c.barker@ucl.ac.uk

Abstract

Background: The well-being of informal carers of people with dementia is an important public health issue. Caring for an elderly relative with dementia may be burdensome and stressful, and can negatively affect the carer's social, family, and professional life. The combination of loss, the physical demands of caregiving, prolonged distress, and biological vulnerabilities of older carers may compromise their physical health, increase social isolation, and increase the risk of anxiety and depressive disorders. Caregiver stress is also linked to negative outcomes for the recipient of care and costs to society, including increased nursing home and hospital admissions. Consequently, carer support interventions are an important component of dementia care. Computer-mediated carer support offers a range of potential advantages compared to traditional face-to-face support groups, including accessibility and the possibility of tailoring to meet individual needs, but there has been little research on its effectiveness so far.

Objective: This mixed-methods study examined the impact of a well-respected UK-based online support forum for carers of people with dementia.

Methods: A total of 61 new forum users completed measures of anxiety (7-item Generalized Anxiety Disorder scale, GAD-7), depression (9-item Patient Health Questionnaire, PHQ-9), and quality of relationship with the person with dementia (Scale for the Quality of the Current Relationship in Caregiving, SQCRC), at baseline and again after 12 weeks of forum usage, within a pre-post design. In addition, 8 participants were interviewed about their experiences with using the forum.

Results: There was an improvement in the quality of the relationship with the person with dementia (SQCRC: P=.003). There was no change in users' depression (PHQ-9) or anxiety (GAD-7) over the 12-week study period. Interview participants reported a range of positive experiences and benefits from using the forum. Limited negative experiences were also reported.

Conclusions: Many of the reported experiences and benefits are unique to online peer support. Further research into online peer support for carers of people with dementia is needed to clarify who benefits under what conditions.

(J Med Internet Res 2014;16(2):e68) doi:10.2196/jmir.3166

KEYWORDS

Alzheimer disease; dementia; caregivers; self-help groups; Internet

Introduction

There are currently about 800,000 people in the United Kingdom with dementia [1], approximately 1% of the total population. Many are cared for at home by a relative or friend, which can negatively affect the carer's social, family, and professional life [2]. The rate of anxiety and depressive disorders is increased in carers [3] and caring may also compromise their physical health [3], mortality [4], and ultimately their ability to function as carers.

Therefore, there is a need to find effective ways of supporting carers of people with dementia. The potential of the Internet for this was recognized early on [5] and there now exists a range of multifaceted interventions with elements of networked support [6]. However, conclusions about effectiveness are difficult to draw because of the varied components within each intervention [7]. One common format is the online forum or online support group (we use these terms interchangeably), which provides a number of potential advantages compared with more traditional support mechanisms. These include logistical advantages of carers being able to access support from their homes and cost advantages to service providers.

The limited research specifically looking at online mutual support for carers of people with dementia has generally focused on the content of messages and posts [5,8]. There has been little research examining the outcome of online mutual support for carers of people with dementia, or research that attempts to understand more about how carers find online mutual support to be helpful.

The present study is a mixed-methods evaluation of a well-respected and well-used UK-based online forum for carers of people with dementia. The quantitative component involved baseline and 12-week follow-up measurements of new forum users' depression, anxiety, and quality of relationship with the care recipient. The qualitative component involved semistructured interviews with new users.

It was hypothesized that after 12 weeks of forum usage: (1) users' anxiety and depression would decrease and (2) the magnitude of this effect would be correlated with the amount of forum usage. Changes in the quality of the relationship with the person with dementia were also examined. No hypotheses were made in relation to this variable because of 2 conflicting possibilities. Although there is some evidence that the relationship quality might deteriorate over time (eg, [9,10]), it might be anticipated that the forum could increase the quality of the relationship. Therefore, the likely effect is that the quality of the relationship will remain roughly stable.

Qualitative interviews examined participants' detailed experiences of being on the forum and possible positive and negative outcomes.

Methods

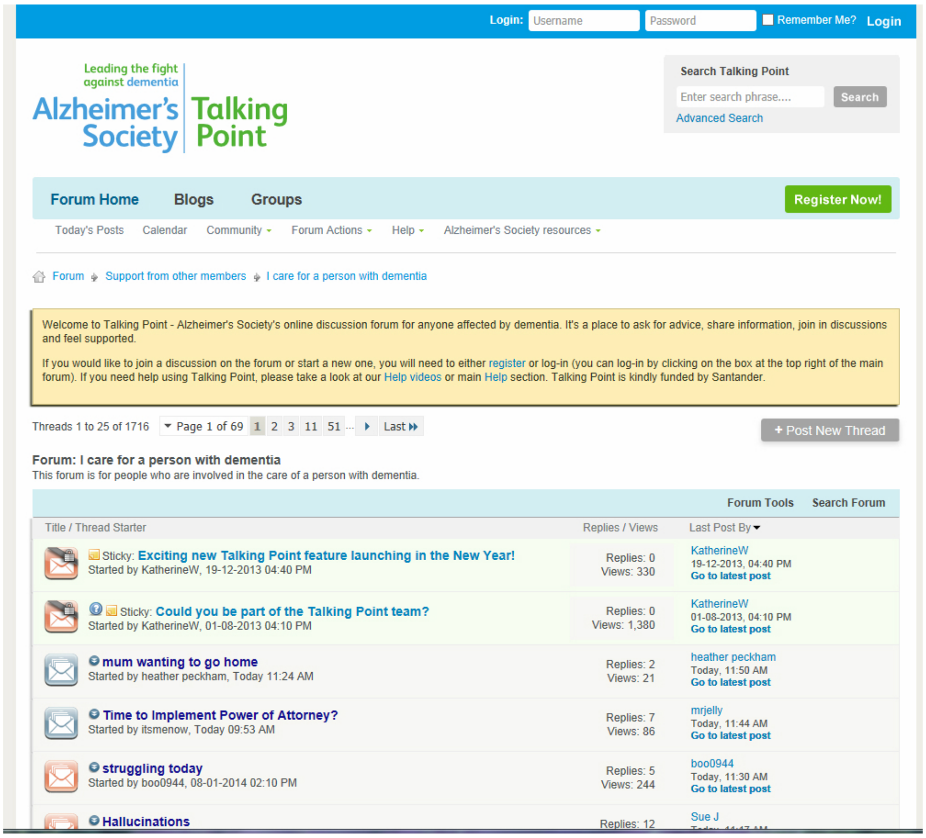
Setting

This research was primarily conducted online, with participants accessing the survey through a link on the forum's home page. There were 8 follow-up face-to-face or telephone interviews. Ethical approval was obtained from the University College London Research Ethics Committee.

Intervention

The forum studied was Talking Point, the UK Alzheimer's Society's online forum for carers of people with dementia [11]. Its home page describes it as "an online support and discussion forum for anyone affected by dementia. It's a place to ask for advice, share information, join in discussions, and feel supported." The forum is well used; on August 20, 2013 there were 873 active members, and at the time of visiting (18:00), there were 444 users online (87 members and 357 guests). It contained a number of different subforums, including "I care for a person with dementia" forum, which was the most active. Figure 1 shows a screenshot of discussion threads on Talking Point.

Figure 1. Screenshot of discussion threads on Talking Point's online forum for carers of people with dementia.



Welcome to Talking Point - Alzheimer's Society's online discussion forum for anyone affected by dementia. It's a place to ask for advice, share information, join in discussions and feel supported.

If you would like to join a discussion on the forum or start a new one, you will need to either register or log-in (you can log-in by clicking on the box at the top right of the main forum). If you need help using Talking Point, please take a look at our Help videos or main Help section. Talking Point is kindly funded by Santander.

Forum: I care for a person with dementia
This forum is for people who are involved in the care of a person with dementia.

Title / Thread Starter	Replies / Views	Last Post By
Sticky: **Exciting new Talking Point feature launching in the New Year!** Started by KatherineW, 19-12-2013 04:40 PM	Replies: 0 Views: 330	KatherineW 19-12-2013, 04:40 PM Go to latest post
Sticky: **Could you be part of the Talking Point team?** Started by KatherineW, 01-08-2013 04:10 PM	Replies: 0 Views: 1,380	KatherineW 01-08-2013, 04:10 PM Go to latest post
mum wanting to go home Started by heather peckham, Today 11:24 AM	Replies: 2 Views: 21	heather peckham Today, 11:50 AM Go to latest post
Time to Implement Power of Attorney? Started by itsmenow, Today 09:53 AM	Replies: 7 Views: 86	mrjelly Today, 11:44 AM Go to latest post
struggling today Started by boo0944, 08-01-2014 02:10 PM	Replies: 5 Views: 244	boo0944 Today, 11:30 AM Go to latest post
Hallucinations	Replies: 12	Sue J

Participants

Eligibility

Inclusion criteria were that participants had to be (1) a new user on their first visit to the forum, (2) an informal carer for a relative or friend with dementia (ie, unpaid), (3) involved in a significant amount of the care of this person, (4) older than 18 years, and (5) fluent in English.

Additionally, participants were eligible for the qualitative interview if (1) they indicated at baseline that they were interested in being interviewed, (2) they completed the survey at 12 weeks, and (3) they visited the forum at least 6 times over the 12-week study period.

Participant Numbers and Response Rate

A total of 128 participants completed the first survey between July 25, 2012 and January 9, 2013. In the 6 months between July 1, 2012 and January 31, 2013, 4177 new users registered on Talking Point; therefore, the percentage of potential participants who took part in the research was low (3.06%). Figure 2 shows numbers of participants at each stage in the research process. Table 1 gives demographic information for the 119 baseline participants.

Figure 2. Participant flow chart.

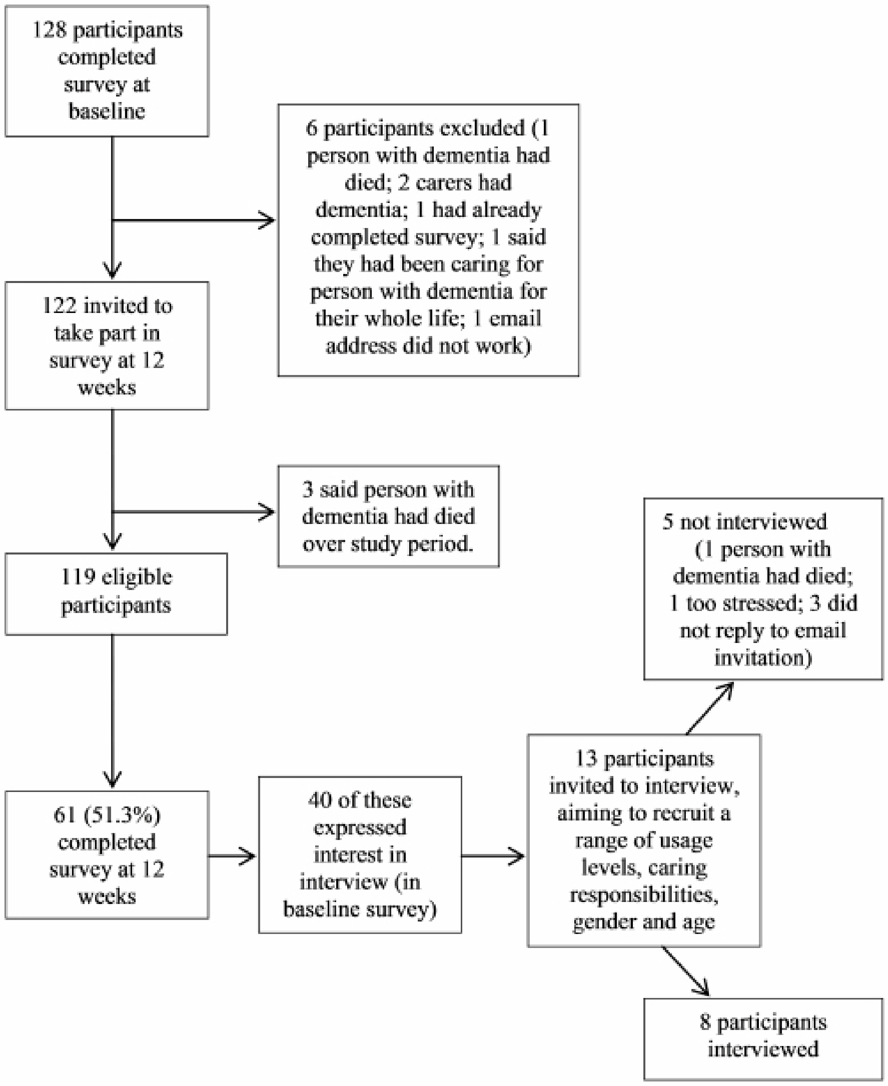


Table 1. Participant information (N=119).

Demographic characteristic	Mean or frequency
Age in years (range 22-86), mean (SD)	56 (11.29)
Number of months caring (range 0-408), mean (SD)	44 (56.94)
Sex, n (%)	
Female	99 (83.2)
Male	18 (15.1)
Missing	2 (1.7)
Ethnicity, n (%)	
White British	112 (94.1)
White other	4 (3.4)
Other	3 (2.4)
Employment status, n (%)	
Employed	58 (48.7)
Unemployed	19 (16.0)
Retired	42 (35.3)
Educational level, n (%)	
Primary school	7 (5.9)
GCSEs/equivalent	22 (18.5)
A levels/equivalent	16 (13.4)
University degree	39 (32.8)
Higher degree/equivalent	18 (15.1)
Other	9 (7.6)
Person being cared for, n (%) [a]	
Father	22 (18.5)
Mother	45 (37.8)
Partner	38 (31.9)
Grandparent	3 (2.5)
Aunt or uncle	3 (2.5)
Sibling	5 (4.2)
Mother- or father-in-law	6 (5.0)
Other	3 (2.5)
Formal support received, n (%)	
General practitioner	24 (20.2)
Mental health worker or counselor	6 (5.0)
Another health or social care professional	27 (22.7)
Memory clinic	22 (18.5)
Other/not specified	8 (6.7)
No formal support	55 (46.2)
Informal support received, n (%)	
Friends and/or family	49 (41.2)
Religious organizations	2 (1.7)
Recreational groups	1 (0.8)
Charities, helplines, or forums	12 (10.1)

Demographic characteristic	Mean or frequency
Other/not specified	5 (4.2)
No informal support	51 (42.8)

[a] Figures total more than 100% because 6 participants reported that they were caring for more than 1 person.

Power Calculation

A recent systematic review [6] found that networked technology interventions for carers of people with dementia had "moderate effects on improving carer stress and depression." Assuming a medium-small effect size (0.35), a sample size of 67 was required for this analysis [12]. The actual sample size achieved was 61, which gave a power of 0.80 at an alpha of .05 to detect an actual effect size of Cohen's *d*=0.36.

Procedure

New users of the forum were invited to take part in the research through an advertisement on the home page, from which they could click on a link to the consent form and survey. An email was also sent to users to alert them to the research, which was also promoted on the Alzheimer's Society's Facebook and Twitter pages. At baseline, participants completed 3 standardized measures: the 9-item Patient Health Questionnaire (PHQ-9), the 7-item Generalized Anxiety Disorder scale (GAD-7), and the Scale for the Quality of the Current Relationship in Caregiving (SQCRC). They also answered demographic questions and questions about their role as a carer.

At 12 weeks after completing the first survey, participants were emailed a link to complete the 3 standardized measures again, as well as questions about their use of the forum over the 12-week period. Those who did not complete the second survey within 2 weeks were sent an email reminder.

At baseline, participants indicated if they were interested in an optional interview about their experiences of using the forum. Interview participants were selected according to inclusion criteria and in order to sample a range of different users, in terms of their sex, age, person they were caring for, and length of time they had been caring.

In accordance with the study risk protocol, participants scoring in the severe range on the PHQ-9 or the GAD-7 were advised via email to contact their general practitioner. A total of 49 participants were sent this email.

Measures

The Patient Health Questionnaire (PHQ-9 [13] is a widely used 9-item measure of depression. Scores of 20 or more suggest severe depression. It has high sensitivity and specificity for diagnosing depression [13], good internal consistency, convergent and discriminant validity, robustness of factor structure, and responsiveness to change [14].

The Generalized Anxiety Disorder scale (GAD-7 [15] is a widely used 7-item measure of anxiety. Scores of 15 or greater indicate severe anxiety [15]. It has good sensitivity and specificity for GAD [15] and is a valid and reliable measure for detecting GAD in the general population [16], as well as social anxiety, panic disorder, and posttraumatic stress disorder [17].

The Scale for the Quality of the Current Relationship in Caregiving 14-item version of the SQCRC (SQCRC-14) [18] asks carers about their relationship with the person that they are caring for, giving equal weight to positive and negative aspects. A higher score implies the presence of warmth and affection and the absence of conflict and criticism in the relationship. The measure has high internal consistency and good face validity [19], but has had little further psychometric investigation.

Qualitative Interview

The qualitative interview asked participants about their experiences using the forum. It covered what they found more and less useful, how they liked to make use of the forum (eg, whether they preferred to write posts or read other people's), and whether they felt that it had made a difference to them and their role as a carer. The rationale was to capture the variety of users' experiences related to both our main outcome variables and also user-defined outcomes. The interview schedule was developed with reference to the literature on peer support, as well as through discussion with the forum's manager and the charity's head of evaluation. The manager and volunteer moderators provided feedback during the development of the interview schedule. Seven interviews were conducted over the telephone and one was face-to-face. The interviews lasted approximately 40 minutes.

Researcher Perspective

It is recommended that the researchers' perspective be disclosed to enhance the credibility of qualitative research [20,21]. The first author, who conducted the interviews and led the analysis, was a white, middle-class, female clinical psychology doctoral student in her twenties. She had no direct experience of caring for a friend or family member with dementia, but had seen the impact of the caring role on other people and was aware of the challenges and stress that this role can bring. The other 2 authors were white, male clinical psychologists, who were generally favorably disposed toward mutual and peer support. All authors attempted to bracket their preconceptions during the analysis.

Analysis

Paired sample *t* tests were used to analyze changes in anxiety from baseline to 12 weeks, and depression and quality of relationship for all participants who completed the survey at 12 weeks. Nonparametric correlations were conducted to examine the relationship between forum usage level and changes in outcome.

Semistructured interviews were transcribed verbatim, with all identifiable data removed to preserve anonymity. They were analyzed using thematic analysis [22], taking an inductive, data-driven approach. Interview transcripts were repeatedly reviewed to become familiar with the data and to ensure that information was represented accurately. During this process, an initial list of ideas was generated. These were grouped into codes and then brought together into meaningful themes, which were then checked against initial codes and the overall dataset, and amended in some instances. Finally, themes were organized into overarching domains. Credibility checks [21,23] involved a third party examining sections of analyzed interview transcripts and providing feedback on codes, themes, and domains.

Respondent validation [20] was used as a further credibility check. Each interviewed participant was emailed a summary of the themes generated from their interview and asked to complete a feedback form. Seven of the 8 interview participants responded to this, either by completing the form or simply by replying to the email, and said that they felt that the list of themes was a good summary of their interview.

The quantitative and qualitative data were analyzed separately, using a concurrent analysis approach [24]. This allows each analysis to stand on its own as an independent perspective on the data, rather than have one influenced by the other.

Results

Quantitative Analysis

Data Preparation

There were no missing data for the 3 main outcome measures at baseline, nor for any participants who completed the survey at 12 weeks. Three participants' usage data was internally inconsistent. For example, they might have reported that they had visited the forum 10 times, but that they had spent 0 minutes on the forum. Where this occurred, all usage data for that participant was coded as missing. Two additional participants gave unclear answers regarding the amount of time spent on the forum, and this was also coded as missing. The distribution of PHQ-9 scores was positively skewed; therefore, a square root transformation was carried out to conduct statistical tests.

Characteristics of Noncompleters and Nonusers

There were no differences in baseline PHQ-9, GAD-7, and SQCRC scores between those who completed the survey at 12 weeks and those who did not (Table 2).

Of the 58 participants whose forum usage data was available, 17 (29%) reported that they had not visited the forum at all over the 12-week study period. There were no differences between baseline scores of users and nonusers on any of the 3 measures (Table 2).

Overall, usage was low (Table 3); 44 (76%) participants visited the forum fewer than 12 times over the 12-week study period. Of those participants who visited the forum, 18 (44%) never replied to any posts, and 20 (49%) never started their own new thread or post.

Table 2. Baseline scores on the Patient Health Questionnaire (PHQ-9), Generalized Anxiety Disorder scale (GAD-7), and the Scale for the Quality of the Current Relationship in Caregiving (SQCRC) for those who completed the survey at 12 weeks (completers) and those who did not (noncompleters) and those who used Talking Point (users) and those who did not (nonusers), and pre-post outcome scores.

Baseline score comparison	Group, mean (SD)		*t* (df)	*P*	Cohen's *d*
	Completers	Noncompleters			
Completed survey at 12 weeks					
PHQ-9	9.75 (6.65)	9.78 (7.35)	0.39 (117)	.69	0.07
GAD-7	10.38 (6.44)	10.38 (6.65)	0.00 (117)	.99	0.00
SQCRC	48.61 (9.20)	46.69 (10.08)	–1.09 (117)	.28	–0.12
Used Talking Point	Users	Nonusers			
PHQ-9	9.98 (6.51)	8.88 (6.78)	0.52 (56)	.61	0.15
GAD-7	9.83 (6.14)	10.71 (6.15)	–0.47 (56)	.64	0.14
SQCRC	49.00 (9.04)	46.88 (9.86)	0.79 (56)	.43	0.23
Pre-post outcome	Baseline	At 12 weeks			
PHQ-9	9.75 (6.65)	9.23 (6.82)	0.75 (60)	.46	0.19
GAD-7	10.38 (6.44)	9.72 (6.63)	0.95 (60)	.35	0.24
SQCRC	48.61 (9.20)	51.36 (9.87)	–3.04 (60)	.003	–0.78

[a] Note that *t* tests were performed on the square root transformed PHQ-9 scores.

Table 3. Reported usage levels on Talking Point over the 12-week study period.

Usage variable	Range	Mean	SD
Number of forum visits	0-200	17.43	37.14
Number of minutes spent on forum	0-3000	373.57	589.86
Number of messages/posts written as a reply to someone else	0-400	14.53	54.76
Number of new threads/posts started	0-50	2.67	8.04

Pre and Post Outcome Comparisons

Of those who completed the survey at 12 weeks, 16 (26%) had a poor relationship with the person with dementia at baseline (scores ≤42 on the SQCRC), 19 (31%) fell in the severe range for anxiety on the GAD-7, and 6 (10%) fell in the severe range for depression on the PHQ-9.

There was a difference between SQCRC scores at baseline and scores at 12 weeks, but not on the PHQ-9 or GAD-7 (Table 2). There is insufficient psychometric data to evaluate the clinical significance of the improvement in SQCGR scores, but according to the scale developers [19], both the pre and post means are in the good range of greater than 42.

To investigate the hypothesis that users’ outcomes are related to their level of forum usage, nonparametric correlations between outcome and usage were examined for the subsample of 40 participants who said that they had visited the forum at least once over the 12-week study period. Only 1 of these 12 correlations was statistically significant; specifically, between total time on the forum and reduction in PHQ scores (Spearman rho=.356, *P*=.03).

Individual Change

Table 4 shows reliable change analysis at the level of individual participants [25] (reliabilities for the 3 measures were obtained from PHQ-9 [26]; GAD-7 [16]; SQCRC [19]). Most participants showed no reliable change.

Table 4. Reliable change analysis for Patient Health Questionnaire (PHQ-9), Generalized Anxiety Disorder scale (GAD-7), and the Scale for the Quality of the Current Relationship in Caregiving (SQCRC).

Measure	Change, n (%)		
	Reliably improved	No change	Reliably deteriorated
PHQ-9	9 (15%)	48 (79%)	4 (7%)
GAD-7	11 (18%)	41 (67%)	9 (15%)
SQCRC	5 (8%)	55 (90%)	1 (2%)

Qualitative Analysis

From the 61 participants who completed the 12-week survey, 40 expressed an interest in being interviewed and 13 were invited (with the aim of having a variety of backgrounds and forum usage). Of these, 8 were able to be interviewed. Table 5 gives their characteristics.

Thematic analysis of the 8 semistructured interviews resulted in 18 themes across 3 domains (Table 6). Each domain is discussed subsequently, and illustrative quotes can be found in Table 6.

Table 5. Interviewee information (n=8).

Participant number	Demographic information	Caring situation
1	Female; age 51 years; white British	Caring for husband (age 53 years) who has frontotemporal dementia. Participant and her husband live together. Caring for approximately 21 months at time of interview.
2	Female; age 48 years; white British	Caring for mother who has a mixed diagnosis of vascular dementia and Alzheimer disease. Mother does not live with participant. Caring for approximately 29 months at time of interview.
3	Female; age 43 years; white British	Caring for father who has dementia and lives locally. Caring for approximately 28 months at time of interview.
4	Male; age 70 years; white British	Caring for wife who has vascular dementia. Participant and his wife live together. Caring for approximately 33 months at time of interview.
5	Male; age 84 years; white British	Caring for wife who has Alzheimer disease. Caring for approximately 30 months at time of interview. Wife moved into care home 1 year previous.
6	Female; age 69 years; white British	Does not consider herself a carer because husband is in the early stages of Alzheimer disease and does not require significant levels of support. At time of interview, it was approximately 29 months since memory problems began.
7	Female; age 63 years; white British	Caring for mother-in-law who lives locally and has dementia. Caring for approximately 12 months at time of interview.
8	Female; age 61 years; white British	Caring for husband who has dementia. Participant and her husband live together. Caring for approximately 18 months at the time of interview.

Table 6. Summary and illustration of domains and themes from the thematic analysis.

Domain and theme	Prevalence[a]	Illustrative quotations
Social similarity		
I am not the only one going through this	Typical	"Every time I realize 'that's me, that's me! I'm going through that, that's me!' I can relate to so much of what's on there." [P1]
Reduced isolation and loneliness	General	"All of a sudden, I'm in the flat on my own. And I have nobody to talk to...so the only thing I've got now, really, is Talking Point" [P5] "Before my son introduced me to that [Talking Point] I felt that I was on my own, even though, like I said, I have fantastic support from the help service and family and everything, it still, I still felt alone." [P1]
Normalizing	Typical	"I'm feeling like I want my dad to die, because I don't want to see him go through this, and he'd be happier. And then you feel full of guilt, but you go on [to Talking Point], and you're not the only one feeling that, or you're not the only one that has felt that. And it makes you feel OK, normal." [P3]
Other users have experience and are therefore understanding	Typical	"And that is the great strength of it. Everybody on Talking Point has hands-on experience of dealing with dementia. And they make allowances for you, as they did for me." [P5]
Being able to share and let off steam	Typical	"If nothing else, it's purely somewhere to let off steam." [P4]
Other users are worse off than I am	Typical	"My goodness, some people are dealing with some incredibly difficult situations...the situation we've got at the moment is not that difficult, we ought to be able to work around it." [P7] "If I'm online and I'm reading other people's posts and it's quite distressing, you know, it's quite easy to get distressed by it. So there are times when I just don't go on it because I think, oh, it's too painful, really, for me." [P2]
Not necessarily needing to post to benefit	General	"I think it's more useful for me to read other people's experiences." [P6]
Being able to give advice and support to other users	General	"It's nice to be giving something back for the information that you're getting. Or being able to support people as the support you receive." [P3]. "It's helpful in that you think, well, at least you know slightly more than someone else...It gives you a little bit of a boost to know that you can help somebody." [P6]
Unique aspects		
I can ask questions and get the support that I might not be able to get, or might not want to get, elsewhere	Typical	"It's not the kind of thing you would read in a leaflet that you pick up anywhere." [P6] "Me personally, I don't want to go sit in front of a—what would you go sit in front of?—a doctor, or a whoever. I don't want to talk that way." [P3]
I can get information and advice that I would not know where else to get	Rare	"For someone who is totally ignorant of Alzheimer's it has answered questions that I don't know where I would have got the answer from." [P6]
Control	Typical	"I can go for months without bothering with it...I'll utilize Talking Point when I need to. That maybe sounds a bit selfish, but I think that's the advantage of it for me." [P2] "You're not seeing what it's going to be like, you're not seeing, you're not hearing what it's, you read what you want to read." [P3] "I find it's very easy to navigate to and from the titles to pick up things that might be useful." [P7]
Anonymity	Typical	"What I say there, I am opening up my heart to a very large extent and I am opening myself up and being totally honest about my feelings, and in some cases I wouldn't want other members of my family, for example, to see some of these things...I wouldn't want them to know exactly how I feel about certain things. Because when you are with other people you never really uncover your true self, do you?" [P5] "I suppose it's when you're sat at home thinking, 'oh crikey'" you know, someone thinks that's really awful. And it puts you in a bit of a panic, but I do think the moderators help because they come back with...solutions." [P2]
Immediate access and responses	Variant	"Now if I have something that is worrying me, I know that I have immediately got someone to talk it over with." [P8]
No time restrictions	Variant	"I think professional support is generally very time limited. And Talking Point isn't time limited." [P2]
Geography is unimportant	Variant	"Particularly with me being mobile around the world as well as the country, it's offered a support that I wouldn't have had otherwise." [P7]
New learning		
Practical learning and information	General	"Off Talking Point, someone said that animals are really good. So I got him a kitten...It was absolutely brilliant, and I thought 'thank God for Talking Point' because when my mum had to go out, he had the kitten to talk to, and the kitten to play with, and the cat even now." [P3]

Domain and theme	Prevalence[a]	Illustrative quotations
Learning how the dementia might progress and what to expect	Typical	"That happened 3 times in the last few weeks, where [husband] hadn't known where he was. And it has been stressful for him. But had I not known about it, I might have worried more. But having read about it [on Talking Point], I thought, mmm, yes, things do happen." [P6] "When you're reading it, you're thinking, 'I've got all this to come,' but at the same time, you're planning." [P3] "I now look for the kind of problems that other people are having, that I am not having...And so that saddens me in a way that I possibly know more and am looking for things, and really I shouldn't be." [P6]
Developing a better understanding of the person with dementia, and consequently becoming a better carer	General	"I don't reason with her any more, I agree, and I can steer the conversation, and I know the kinds of things to say and the kinds of things to stay away from. And I think I'm a much better person for Talking Point. A much better companion for my wife, I know that." [P5]

[a] General: theme applies to all or all but one of the participants (7-8); typical: theme applies to more than half of the participants (5-6); variant: theme applies to up to half of participants (3-4); rare: theme applies to 1 or 2 participants (1-2).

Domain 1: Social Similarity

All interview participants emphasized the importance of the forum being for people who are in the same situation—caring for someone with dementia. For the most part, this was considered to be a great benefit, although some also noted disadvantages. Six participants spoke about connecting with other carers on the forum, and their comfort in knowing that "I am not the only one going through this." Seven participants noted that being able to use the forum reduced their isolation and loneliness. Some had experienced isolation as a result of their caring role, whereas most participants who felt isolated and alone had social support but still felt alone before joining the forum. Six participants noted the normalizing effect of the forum. Six participants reflected that because other users had experience with dementia they were understanding about the struggles that forum users face. This also meant that background information did not need to be explained to others on the forum. Five participants reflected on the value of the forum in being able to "share and let off steam" with others. Several participants commented that other users of the forum were in a worse situation than they were, which made 3 participants feel more positive about their own situation. However, for 3 interviewees, reading other users' posts had been distressing at times. Seven participants noted that because everyone on the forum was a carer of someone with dementia, they did not necessarily need to post to benefit, and that simply reading the posts was useful. Seven said that they had been able to give advice and support to other users of the forum. For some, this helped people to feel that they were giving back to the forum. For others, it also showed them that they had valuable knowledge to pass on.

Domain 2: Unique Aspects

All 8 participants compared the forum to other sources of information and support. Sometimes this was explicitly discussed, but more often it was implied by the advantages the forum offered them. Six participants commented that they could ask questions and get the support that they might not be able to get elsewhere or that they might not want to get elsewhere. Two commented that they would not know where else to get information and advice. Five participants reflected that through using the forum they had control, including control over frequency of usage and, to a certain extent, the ability to avoid posts that were too upsetting. Many also discussed only reading posts that were of interest or personally relevant. Four participants commented on the benefits of anonymity on the forum, including being able to both be more open and honest and to discuss problems that might be uncomfortable to discuss face-to-face. Four participants noted that they had seen inappropriate or judgmental posts on the forum, which may be another consequence of the forum's anonymity. One participant (P2) had received replies from other users who "put something on about how disgusting it [a post she had made] was." She and others reflected that the forum's volunteer moderators were generally helpful in these situations. Three participants reflected that the forum allowed immediate access and response. Three others commented that it could be accessed 24-hours a day for as long as need be. Three participants noted that geography was unimportant. Two participants had lived or currently lived outside the United Kingdom and both reflected on the benefits of the forum for them.

Domain 3: New Learning

All participants described learning new information, and many said that what they had learned on the forum had helped them to become better carers. All 8 participants commented that the forum had provided them with practical advice and information. This ranged from information about Attendance Allowance (a government benefit for people with disabilities), to information about how to deal with people with dementia turning on an empty microwave, to one participant buying her father a cat. Five participants noted that they learned about how the dementia might progress and what to expect. For some, this meant that when certain events did occur, they were more prepared for them and less anxious when they did occur. For others, this information meant that they felt forewarned and, therefore, forearmed in terms of the next stages of the dementia. One participant expressed sadness that through what she had read on the forum, she had started to look for certain behaviors in her husband. Seven participants reflected that they developed a better understanding of the person with dementia and consequently became a better carer.

Discussion

Principal Findings

There was an improvement in the quality of the relationship with the person with dementia. There was no change in users' depression or anxiety over the 12-week study period in contrast to other studies of the outcome of online support [27]. Interview participants reported a range of positive experiences and benefits from using the forum. Limited negative experiences were also reported.

The hypotheses that after 12 weeks of forum usage, users' anxiety and depression would decrease were not supported, although depression and anxiety levels did not increase. This may indicate no effect of forum usage. However, interpretation is dependent upon understanding of the normal expected trajectory of carer mood, and this is complicated because models predict anything from deteriorating to increasing psychological health over time [28]. Additionally, the progression of psychological distress in carers is affected by a range of factors, including intrinsic variables, such as carer gender [29], and more contextual factors, such as the relationship between carer and care recipient [30], which varied across the sample. It may also be that participants' low usage of the forum accounts for these findings because there is evidence of a dose-response relationship in online forums for mental health conditions [31].

Qualitative results provide some insight into how the forum may have improved the carer's relationship with the person with dementia; for example, through carers learning more about how to interact with someone with dementia and feeling as though the information that they had gathered helped them to become a better carer. Many participants' forum usage levels were low, and it is possible that being aware of the forum as a resource that was available should they wish to use it was enough for some participants to improve the quality of their relationship with the person that they were caring for. In addition, it is possible that those who signed up to the forum were individuals who had made the decision to learn and develop as much as possible in their role as a carer, and these individuals were motivated to have a good quality relationship with the person with dementia. It is also possible, however, that the improvement in scores on this measure simply represents regression to the mean.

For all 3 measures, most participants' scores neither reliably improved nor reliably deteriorated. More participants reliably improved than reliably deteriorated. Although the present study only covered a 12-week period, this finding is encouraging given that relationship quality is likely to deteriorate over time [9,10]. Research into the trajectories of anxiety and depression is more mixed [28], but this is also an encouraging finding in relation to these measures.

The 8 carers interviewed were generally very positive. Their reasons for using the forum echo previous research, which found that the primary functions of such groups are to exchange information, connect to others, and to obtain emotional support [32].

For interviewees, the fact that other forum users were in a similar situation to them had a range of benefits. Carers of people with dementia frequently report feelings of isolation and inadequate social support [33], but interview participants reported that the perceived similarity of other users reduced isolation and loneliness, allowed them to "let off steam," and feel more normal and understood. These experiences, which correspond to Yalom and Leszcz's [34] therapeutic factors in group therapy (universality, altruism, guidance, imparting information, and catharsis) are similar to those reported in a study of mental health service users' information needs [35].

Several interviewees noted that seeing that others were worse off helped them to feel better about their own situation. No interviewee mentioned comparing themselves to individuals better off than them. Benefits of downward comparison have been seen found in online support groups for heath conditions [36,37]. However, for several participants, reading the stories of other users who were in worse situations was distressing.

Interviewees reported deriving benefit from helping others, in-line with the helper therapy principle [38], which is unique to peer support. Providing help can increase feelings of competence, equality, social usefulness, independence, and social value, and allows individuals to view themselves as having strengths as well as needs [39]. The reported benefits of lurking (using the forum without actually posting) are echoed in another study [40] that found lurkers on an online support group reported a range of benefits indicating that reading messages may be as beneficial as interacting with the group.

Several interviewees reflected that the type of information and support available on the forum was either not available to them elsewhere or that they would not wish to obtain it elsewhere. The forum may either work as a complement to other services, or offer a service to people who do not access other services. A number of interviewees were not the main carer of the person with dementia; therefore, perhaps, more traditional services are not available to them.

The forum's anonymity provided some interviewees with the freedom to say more than they would be able to otherwise. This online disinhibition effect [41] brings the advantage that users may be more open about their feelings allowing access to emotional support. However, several interview participants also considered some posts to be judgmental or inappropriate. Helpful moderators (as found on Talking Point) are important to ameliorate this more toxic online disinhibition effect [41]. Additionally, some participants said that the forum allowed them to read only the posts that they wanted to, allowing them to avoid distressing posts.

Several participants noted that reading posts helped them to learn more about progression of dementia and to be better prepared for future problems. However, others felt that some posts were too distressing to read. With a disease that involves worsening of symptoms over time, more information and less uncertainty about the future can provide relief, but it can also cause distress and more information is not always positive. This tension has also been found in support groups for people with motor neuron disease and their carers [36].

Participants also noted that the forum taught them more about how to interact with someone with dementia and to become a better carer or companion. Consequently, the forum may have benefits for the person with dementia through improved care.

There is an apparent discrepancy between the positive qualitative experiences of the site and the modest pre-post changes. This may be because of the contrast between psychological outcome variables measured quantitatively and the qualitatively reported benefits, which were largely related to general therapeutic group factors [34] as well as to practical advice and learning. This raises questions about what carers hope to gain from the site and provides useful information about outcome measures for future studies, a contentious issue in this field [42,43].

Limitations

There were several limitations to the study. There was a low usage rate that affected the ability to reliably assess the impact of the forum and skewed the data. However, this may reflect real-life usage and, as a number of participants noted, a benefit was being able to visit the forum as little or as often as desired. Additionally, inclusion criteria for interviewees excluded those who used the forum less or disengaged from using the forum. Their views may have differed from those who were interviewed, but those interviewed still provided useful information about people who do engage with the forum, of whom there are many. Only 8 participants were interviewed, which is a relatively small number, and it is possible that a larger sample of interviewees would have provided different or additional data. Whether these 8 interviewees' experiences reflect the experiences of the wider group of forum users is difficult to ascertain.

Participants received various levels of additional formal and informal support in their roles as carers, which may further complicate and limit the scope of conclusions that can be made about the specific role of the forum. In addition, only approximately 3% of possible participants took part in the research. These sampling issues inevitably limit the extent of conclusions that can be made, and mean that findings from the quantitative data, in particular, are tentative. The lack of a control group in the present research means that these outcomes cannot be compared to a group of carers who did not have access to Talking Point; therefore, findings cannot necessarily be attributed to forum usage as opposed to other factors. Future studies would benefit from using a randomized controlled design.

Finally, the caregivers in our sample tended to be late in their caring trajectory (adult child vs spouse caregivers) with the median age suggesting likely sandwich generation carers. These caregiver variables have systematic effects on caregiver mental health outcomes [44]; the number of participants scoring in the severe range for depression was relatively low [45] although this was not the case for anxiety [46]. Given these sample characteristics and evidence of low depression rates, caution is needed when generalizing the results to the wider population of caregivers.

Implications

The qualitative data indicated unique benefits from peer support/group therapy, such as not feeling alone and feeling understood through shared experience. Some of these benefits are unique to online support, such as being continually and flexibly accessible as well as enabling honesty within an anonymous online environment. Therefore, clinicians may wish to direct carers to such online peer support forums. In an economic climate in which services are often being reduced, online peer support is likely to become more prevalent and may be the only support that some carers receive. It is important that this area continues to be researched so that carers can derive maximum benefit from online peer support forums.

Future research could investigate further what outcomes are important to carers and specifically consider these outcomes in the evaluation of online carer support interventions. In addition, more research needs to be conducted into whether different types of carers (eg, according to gender and ethnicity) derive different types or levels of benefit from online peer support forums because previous research (eg, [47,48]) suggests that interventions do not affect carers uniformly. In order that carers receive the best support possible, it would also be worthwhile investigating whether there are certain types of support that work well when offered in conjunction with online peer support and vice versa. Interventions for carers of people with dementia are often most effective when they are multicomponent in nature [49,50], although this makes evaluation and attribution of any observed effects more complex [50].

Acknowledgments

We would like to thank the carers of people with dementia who took part in this research and the Alzheimer's Society and staff of Talking Point, particularly Katherine Watt and Linda Sheldrake, for their support.

Conflicts of Interest

None declared.

References

1. Alzheimer's Society. Statistics URL: http://www.alzheimers.org.uk/statistics [accessed 2013-11-21] [WebCite Cache ID 6LIVnScwz]
2. Schulz R, O'Brien AT, Bookwala J, Fleissner K. Psychiatric and physical morbidity effects of dementia caregiving: prevalence, correlates, and causes. Gerontologist 1995 Dec;35(6):771-791. [Medline: 8557205]
3. Schulz R, Martire LM. Family caregiving of persons with dementia: prevalence, health effects, and support strategies. Am J Geriatr Psychiatry 2004;12(3):240-249. [Medline: 15126224]
4. Schulz R, Beach SR. Caregiving as a risk factor for mortality: the Caregiver Health Effects Study. JAMA 1999 Dec 15;282(23):2215-2219. [Medline: 10605972]

5. Brennan PF, Moore SM, Smyth KA. The effects of a special computer network on caregivers of persons with Alzheimer's disease. Nurs Res 1995;44(3):166-172. [Medline: 7761293]
6. Powell J, Chiu T, Eysenbach G. A systematic review of networked technologies supporting carers of people with dementia. J Telemed Telecare 2008;14(3):154-156. [doi: 10.1258/jtt.2008.003018] [Medline: 18430288]
7. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
8. White MH, Dorman SM. Online support for caregivers. Analysis of an Internet Alzheimer mailgroup. Comput Nurs 2000;18(4):168-76; quiz 177. [Medline: 10939186]
9. Gallagher-Thompson D, Dal Canto PG, Jacob T, Thompson LW. A comparison of marital interaction patterns between couples in which the husband does or does not have Alzheimer's disease. J Gerontol B Psychol Sci Soc Sci 2001 May;56(3):S140-S150. [Medline: 11316839]
10. de Vugt ME, Stevens F, Aalten P, Lousberg R, Jaspers N, Winkens I, et al. Behavioural disturbances in dementia patients and quality of the marital relationship. Int J Geriatr Psychiatry 2003 Feb;18(2):149-154. [doi: 10.1002/gps.807] [Medline: 12571824]
11. Alzheimer’s Society. Talking Point Forum home page URL: http://forum.alzheimers.org.uk/forum.php [accessed 2013-12-02] [WebCite Cache ID 6LZRcQjWK]
12. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
13. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
14. Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br J Gen Pract 2008 Jan;58(546):32-36 [FREE Full text] [doi: 10.3399/bjgp08X263794] [Medline: 18186994]
15. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
16. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
17. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 6;146(5):317-325. [Medline: 17339617]
18. Spruytte N, Van Audenhove C, Lammertyn F. Internal Report: The Scale for the Quality of the Current Relationship. Leuven: LUCAS-KULeuven; 2000.
19. Spruytte N, Van Audenhove C, Lammertyn F, Storms G. The quality of the caregiving relationship in informal care for older adults with dementia and chronic psychiatric patients. Psychol Psychother 2002 Sep;75(Pt 3):295-311. [Medline: 12396755]
20. Barker C, Pistrang N. Quality criteria under methodological pluralism: implications for conducting and evaluating research. Am J Community Psychol 2005 Jun;35(3-4):201-212. [doi: 10.1007/s10464-005-3398-y] [Medline: 15909795]
21. Elliott R, Fischer CT, Rennie DL. Evolving guidelines for publication of qualitative research studies in psychology and related fields. Br J Clin Psychol 1999 Sep;38 (Pt 3):215-229. [Medline: 10532145]
22. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
23. Stiles WB. Quality control in qualitative research. Clinical Psychology Review 1993 Jan;13(6):593-618. [doi: 10.1016/0272-7358(93)90048-Q]
24. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research. 2nd edition. London: Sage Publications, Inc; 2010.
25. Evans C, Margison F, Barkham M. The contribution of reliable and clinically significant change methods to evidence-based mental health. Evidence-Based Mental Health 1998 Aug 01;1(3):70-72. [doi: 10.1136/ebmh.1.3.70]
26. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28(1):71-77. [doi: 10.1016/j.genhosppsych.2005.07.003] [Medline: 16377369]
27. 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]
28. Pot AM. The care-giving stress process. In: Jones GMM, Miesen BML, editors. Care-Giving in Dementia: Research and Applications. 3rd ed. Hove: Brunner-Routledge; 2004.
29. Schulz R, Williamson GM. A 2-year longitudinal study of depression among Alzheimer's caregivers. Psychol Aging 1991 Dec;6(4):569-578. [Medline: 1777145]
30. Zarit SH, Whitlatch CJ. Institutional placement: phases of the transition. Gerontologist 1992 Oct;32(5):665-672. [Medline: 1427279]

31. 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]
32. Mo PK, Coulson NS. Developing a model for online support group use, empowering processes and psychosocial outcomes for individuals living with HIV/AIDS. Psychol Health 2012;27(4):445-459. [doi: 10.1080/08870446.2011.592981] [Medline: 21854088]
33. Stoltz P, Udén G, Willman A. Support for family carers who care for an elderly person at home - a systematic literature review. Scand J Caring Sci 2004 Jun;18(2):111-119. [doi: 10.1111/j.1471-6712.2004.00269.x] [Medline: 15147473]
34. Yalom ID, Leszcz M. The Theory and Practice of Group Psychotherapy. New York: Basic Books; 2005.
35. Powell J, Clarke A. Information in mental health: qualitative study of mental health service users. Health Expect 2006 Dec;9(4):359-365. [doi: 10.1111/j.1369-7625.2006.00403.x] [Medline: 17083562]
36. Locock L, Brown JB. 'All in the same boat'? Patient and carer attitudes to peer support and social comparison in Motor Neurone Disease (MND). Soc Sci Med 2010 Oct;71(8):1498-1505. [doi: 10.1016/j.socscimed.2010.06.043] [Medline: 20719422]
37. 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]
38. Riessman F. Ten self-help principles. Soc Policy 1997;28:6-19.
39. Roberts LJ, Salem D, Rappaport J, Toro PA, Luke DA, Seidman E. Giving and receiving help: interpersonal transactions in mutual-help meetings and psychosocial adjustment of members. Am J Community Psychol 1999 Dec;27(6):841-868. [Medline: 10723537]
40. Malik SH, Coulson NS. A comparison of lurkers and posters within infertility online support groups. Comput Inform Nurs 2011 Oct;29(10):564-573. [doi: 10.1097/NCN.0b013e318224b31d] [Medline: 21709546]
41. Suler J. The online disinhibition effect. Cyberpsychol Behav 2004 Jun;7(3):321-326. [doi: 10.1089/1094931041291295] [Medline: 15257832]
42. Magnusson L, Hanson E, Nolan M. The impact of information and communication technology on family carers of older people and professionals in Sweden. Ageing and Society 2005 Aug 23;25(05):693-713. [doi: 10.1017/S0144686X05003673]
43. Schulz R, O'Brien A, Czaja S, Ory M, Norris R, Martire LM, et al. Dementia caregiver intervention research: in search of clinical significance. Gerontologist 2002 Oct;42(5):589-602 [FREE Full text] [Medline: 12351794]
44. Pinquart M, Sörensen S. Spouses, adult children, and children-in-law as caregivers of older adults: a meta-analytic comparison. Psychol Aging 2011 Mar;26(1):1-14. [doi: 10.1037/a0021863] [Medline: 21417538]
45. Cuijpers P. Depressive disorders in caregivers of dementia patients: a systematic review. Aging Ment Health 2005 Jul;9(4):325-330. [doi: 10.1080/13607860500090078] [Medline: 16019288]
46. Cooper C, Balamurali TB, Livingston G. A systematic review of the prevalence and covariates of anxiety in caregivers of people with dementia. Int Psychogeriatr 2007 Apr;19(2):175-195. [doi: 10.1017/S1041610206004297] [Medline: 17005068]
47. Eisdorfer C, Czaja SJ, Loewenstein DA, Rubert MP, Argüelles S, Mitrani VB, et al. The effect of a family therapy and technology-based intervention on caregiver depression. Gerontologist 2003 Aug;43(4):521-531 [FREE Full text] [Medline: 12937331]
48. Mahoney DF, Tarlow BJ, Jones RN. Effects of an automated telephone support system on caregiver burden and anxiety: findings from the REACH for TLC intervention study. Gerontologist 2003 Aug;43(4):556-567. [Medline: 12937334]
49. Beauchamp N, Irvine AB, Seeley J, Johnson B. Worksite-based internet multimedia program for family caregivers of persons with dementia. Gerontologist 2005 Dec;45(6):793-801. [Medline: 16326661]
50. Pillemer K, Suitor JJ. Peer support for Alzheimer's caregivers: Is it enough to make a difference? Research on Aging 2002 Mar 01;24(2):171-192. [doi: 10.1177/0164027502242001]

Abbreviations

GAD-7: 7-item Generalized Anxiety Disorder scale
PHQ-9: 9-item Patient Health Questionnaire
SQCRC: Scale for the Quality of the Current Relationship in Caregiving

Edited by G Eysenbach; submitted 05.12.13; peer-reviewed by J Powell, M Bastawrous; comments to author 23.12.13; revised version received 23.01.14; accepted 25.01.14; published 28.02.14.

Please cite as:
McKechnie V, Barker C, Stott J
The Effectiveness of an Internet Support Forum for Carers of People With Dementia: A Pre-Post Cohort Study
J Med Internet Res 2014;16(2):e68
URL: http://www.jmir.org/2014/2/e68/
doi:10.2196/jmir.3166
PMID:24583789

Original Paper

Smoking Behavior and Use of Tobacco Industry Sponsored Websites Among Medical Students and Young Physicians in Argentina

M Victoria Salgado[1], MD; Raul Mejia[2], MD, PhD; Celia P Kaplan[3], MA, DrPH; Eliseo J Perez-Stable[3], MD

[1]Universidad de Buenos Aires, Buenos Aires, Argentina
[2]Centro de Estudios de Estado y Sociedad, Buenos Aires, Argentina
[3]University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Eliseo J Perez-Stable, MD
University of California, San Francisco
3333 California St
Suite 335
San Francisco, CA, 94118
United States
Phone: 1 415 514 8662
Fax: 1 415 502 8291
Email: eliseops@medicine.ucsf.edu

Abstract

Background: Internet-based marketing has become an attractive option for promoting tobacco products due to its potential to avoid advertising restrictions. In Argentina, several cigarette brands have designed websites for the local market, which promote user participation.

Objective: The intent of the study was to report on the use of tobacco company-sponsored websites by medical students and recently graduated physicians.

Methods: An online self-administered survey was conducted among eligible medical students and recent graduates from the University of Buenos Aires (UBA). Sampling was from lists of email addresses of students enrolled in two required courses. Eligibility criteria were ages 18-30 years and reporting on smoking status. Questions on Internet use included accessing a tobacco brand website at least once during their lifetime and any use of tobacco promotional materials.

Results: The response rate was 35.08% (1743/4969). The final sample included 1659 participants: 73.06% (1212/1659) were women and mean age was 26.6 years (SD 1.9). The majority were current medical students (55.70%, 924/1659) and 27.31% (453/1659) were current smokers. Men were more likely to report having seen a tobacco advertisement on the Internet (*P*=.001), to have received a tobacco promotion personally addressed to them (*P*=.03), to have used that promotion (*P*=.02), and to have accessed a tobacco-sponsored website (*P*=.01). Among respondents, 19.35% (321/1659) reported having accessed a tobacco-sponsored website at least once in their lifetime and almost all of them (93.8%, 301/321) accessed these sites only when it was necessary for participating in a marketing promotion. Most people logging on for promotions reported entering once a month or less (58.9%, 189/321), while 25.5% (82/321) reported accessing the tobacco industry Internet sites once a week or more. In adjusted logistic regression models, participants were more likely to have accessed a tobacco brand website if they were former smokers (OR 2.45, 95% CI 1.42-4.22) or current (OR 8.12, 95% CI 4.66-14.16), if they reported having seen a tobacco advertisement on the Internet (OR 2.44, 95% CI 1.77-3.37), received a tobacco promotion personally addressed to them (OR 5.62; 95% CI 4.19-7.55), or used one of these promotions (OR 14.05, 95% CI 9.21-21.43). Respondents were more likely to be current smokers if they received a tobacco promotion (OR 2.64, 95% CI 2.02-3.45) or if they used one of these promotions (OR 1.93, 95% CI 1.31-2.85).

Conclusions: Our study suggests that tobacco industry websites reach medical students and young physicians in a middle-income country with their marketing promotions. Current or proposed legislation to ban tobacco advertising needs to include Internet sites and related social media.

(J Med Internet Res 2014;16(2):e35) doi:10.2196/jmir.2528

KEYWORDS

smoking; medical students; tobacco websites; Latin America

Introduction

Tobacco use constitutes the leading global cause of preventable death with approximately one death every six seconds, accounting for 10% of all adult deaths [1]. There are at least one billion smokers in the world and 80% of them live in low- and middle-income countries [1]. Argentina, a South American middle-income country, has a prevalence of tobacco smoking of 27.1% [2]. Tobacco industry investment in marketing and advertising in Argentina reached approximately US $18 million in 2002. Although Argentina's president signed the World Health Organization (WHO) Framework Convention on Tobacco Control in 2003, ratification has been delayed so far by the active lobbying of regional legislators from tobacco-growing provinces influenced by tobacco industry positions [3].

The Internet is now one of the most widely used communication channels in the world. Although many countries have restricted advertising of harmful products such as tobacco, Internet-based marketing has the potential to avoid such restrictions [4] while reaching a large number of people. Thus, Internet advertising has become an attractive option for promoting tobacco products and the rise in use of social networking and user-generated content websites is increasing product promotion through electronic media. In a study from 2010, more than 500 Facebook pages were related to British American Tobacco products [5]. Studies that analyzed smoking content on the site YouTube reported the presence of several pro-tobacco videos [6,7], consistent with indirect marketing activity by tobacco companies or their proxies [6]. In Argentina, almost 5 million homes have broadband Internet access covering at least 55% of the population [8,9]. Several cigarette brands have designed websites for the local market that promote user participation whenever possible and target specific products. For example, it is not unusual for a person reached by a marketing promotion to be required to go online in order to participate in a lottery or to obtain a gift, and such promotions usually come included in cigarette packages. Depending on the specific website, content includes online chatting and online gaming, and allows users to provide brand marketing ideas and to participate in marketing campaigns. For example, RJ Reynolds designed a strategy where consumers could comment on the packaging design of four cigarette flavors on the Internet [10].

Health professionals have the potential to play a fundamental role in tobacco control strategies. The WHO has called on health professional organizations to encourage and support their members to be role models by not using tobacco products and by promoting a tobacco-free culture [11]. It has been shown that physicians' smoking status can impact their interaction with patients and may influence their cessation counseling practices [12-14]. In Argentina, smoking prevalence among physicians to date is similar to that of the general population, unlike what has happened in the United States [15,16]. As new cohorts of physicians enter the medical profession, their smoking behaviors and attitudes will influence tobacco prevention and control efforts. In this context, current medical students and recent graduates constitute an important study population, given their potential impact on the health care system and the population's health and the fact that these students have grown up with a common exposure to generalized use of new information technologies. Several survey studies around the world have estimated the prevalence of tobacco smoking among medical students and reported rates as low as 2.8% in Uganda and as high as 43.3% in Albania [17]. This variation can also be observed among high-income countries, with a smoking rate of 3.8% among medical students in Wales [18], 25.2% in Berlin, Germany [19], and 20.8% among 350 health professional students in Texas [20]. A survey conducted in 2005 among a sample of third-year medical students in Buenos Aires, Argentina, showed that 35.5% (33.4% of men and 36.5% of women) were current smokers [17]. This prevalence was one of the highest in South America compared to smoking rates reported from Brazil (19.5% of men and 14.6% of women) and Chile (27.1% of men and 29.3% of women) [21].

The relationship between Internet marketing strategies by the tobacco industry and its effect on young adults needs further study. Given that medical students in Argentina smoke at similar rates as the general population and they constitute an especially important target population in implementing national tobacco control strategies, we considered that they would be an appropriate group in which to evaluate the effect of tobacco industry marketing strategies. This study reports on a survey of the use of tobacco industry-sponsored websites among a population of current and recent medical students in Buenos Aires, Argentina.

Methods

Setting, Sampling, and Procedures

In 2011, a survey was conducted among current and former medical students in Ciudad de Buenos Aires, Argentina and this analysis focused on the reported use of tobacco industry-sponsored websites. Buenos Aires is the largest city in Argentina and the school of medicine at the University of Buenos Aires (UBA) has over 15,000 students enrolled in the 6-year program [22].

Participants' email addresses were obtained from the list of students enrolled in the required one-year rotating internship that constitutes the sixth year of medical school, from the Pharmacology Department of the UBA School of Medicine and the largest teaching hospital affiliated with UBA (Hospital de Clínicas Jose de San Martin). This survey was conducted between March and May 2011. The Institutional Review committee of the Hospital de Clínicas, UBA, Argentina, approved the protocol.

Potential participants were invited via email to complete an online, self-administered Spanish-language survey regarding tobacco smoking behaviors and their beliefs and attitudes toward tobacco industry marketing. Each email message provided a unique link to the survey in order to prevent multiple entries from the same email account. Participants initially had to complete a consent form, where they were informed about the study objectives, the length of the instrument, and were reassured about the voluntary and confidential character of the survey. To be eligible, a person had to be between 18 and 30 years old, a former or current medical student at UBA, and had to report on their smoking status. In addition to the initial email invitation to participate, an additional 14 reminder messages were sent over a period of 2-3 months. A lottery for theater tickets was used as an incentive. We attempted to contact all persons on the lists obtained. The survey was conducted using DatStat Illume software. Data was password protected and stored with limited access to research team members only.

Development of Survey Items

The survey included questions regarding tobacco use, intention to quit, knowledge of tobacco's health effects, and training in helping patients quit. These items were adapted from the Spanish-language version of the Global Health Professionals Survey (GHPS), a survey of third-year students attending dental, medical, nursing, or pharmacy schools [23], from a questionnaire previously developed for a study of youth in Jujuy, Argentina [24], and from a survey developed by the research team for a cessation study in clinical settings. The smoking media literacy (SML) scale that was developed in the United States and adapted for use in Argentina and is used to assess smoking media literacy and its relationship to smoking history among high school students was included [25,26]. English items underwent forward-backward translation from English to Spanish and were reconciled by three native Spanish-speaking Argentinean staff.

We developed additional items regarding respondents' access to tobacco brand websites and the use of their marketing promotions. Linguistic and content validation was conducted through a series of semi-structured interviews with medical students. The research team revised the items for a final version of the questionnaire. Pilot testing of the instrument was conducted with resident physicians recently graduated from medical school evaluating situational factors and time of administration. Instructions were developed to address the importance of providing accurate answers and the confidential nature of the survey.

Age, gender, and student status (current medical student or recently graduated physician) were asked. Due to the association with smoking [24,27], screening for depression symptoms was ascertained by asking whether the respondent had felt sad or blue or had felt little interest or pleasure in doing things in the previous month [28,29]. An affirmative answer to either of these questions was considered positive screening for depression.

Use of Internet and Tobacco Brand Websites

Participants were asked about how frequently they used the Internet, whether they remembered having seen a cigarette advertisement on the Internet, if they have ever received a personally addressed marketing promotion or free tickets to a cultural event from a tobacco company with a specific brand, and if yes, if they ever used it. "Access to a tobacco website" was defined as having visited an institutional or promotional cigarette brand website at least once in their lifetime. Among participants who accessed a tobacco website, we asked if they only accessed them when the cigarette box included a marketing promotion that required accessing a website to obtain a free gift or to participate in a lottery or if they accessed the sites even if they had no intention of participating in the marketing promotion. In each case, we inquired about the frequency of accessing those sites ("from once a day or more" to "once a month or less").

Smoking Behavior

Respondents were considered "ever smokers" if they tried at least one cigarette puff in their lifetime and "never smokers" if they had never tried even a puff. Current smokers reported smoking at least one puff in the 30 days prior to the survey. Among current smokers, a "daily smoker" smoked at least one cigarette each day of the last 30, while a "non-daily smoker" (still a current smoker) did not smoke every day. A "former smoker" is a person that had tried a cigarette at least once but had not smoked in the previous 30 days.

Data Analysis

Descriptive analysis compared the sample by gender and smoking status. Statistical significance was defined with a two-sided alpha of .05. Using multivariate logistic regression, we constructed different models where the outcomes were whether the participant ever accessed a tobacco brand website or was a current smoker. Models were adjusted for age, gender, student status, presence of depression, daily use of Internet, had seen tobacco advertisements on the Internet, received a tobacco marketing promotion addressed to them, and used a tobacco marketing promotion addressed to them. The model where the outcome was "access a tobacco brand website" was also adjusted for "smoking status". The model where the outcome was "current smoker" was also adjusted for "access a tobacco brand website". In these models, each variable was adjusted for all the others.

Results

Overview

Of 4969 unique names and email addresses included in the databases obtained, 35.08% (1743/4969) completed the survey. Among these 1743 respondents, 84 were excluded from the sample because they had never been in medical school (n=28), were older than 30 years (n=48), or they did not report on their smoking status (n=8). The final sample for analysis included 1659 current or recently graduated medical students. Because some participants did not complete the whole survey and some questions were only shown conditioned on a previous response, the number of answers for a specific question may be less than the total number of participants.

Demographics and Exposure to Tobacco Marketing

Table 1 summarizes the characteristics of the sample and reported use of Internet outcomes. Among the 1659 participants, 73.06% (1212/1659) were women, 55.70% (924/1659) were current students, 43.58% (723/1659) were physicians, and 0.72% (12/1659) had dropped out of medical school by the time of the survey. Close to one-third of the participants (30.32%, 503/1659) were between 20 and 25 years, 37.25% (618/1659) were between 26 and 27 years, and 32.43% (538/1659) were between 28 and 30 years old. The mean age was 26.6 years (SD 1.9). Of the respondents, 26.88% (446/1659) reported remembering seeing a tobacco advertisement on the Internet but a majority (51.05%, 847/1659) did not recall. A tobacco marketing promotion specifically addressed to the participant was received by 29.05% (482/1659), and 10.07% (167/1659) used at least one of these promotions (Table 1).

Women were more likely to report positive screening for depression (55.12% vs 45.19%; *P*<.001). Men were more likely to report having seen a tobacco advertisement on the Internet (31.77% vs 25.08%; *P*=.001), to have received a tobacco marketing promotion personally addressed to them (31.32% vs 28.22%; *P*=.03), to have used that marketing promotion offer (11.86% vs 9.41%; *P*=.02), and to have accessed a tobacco-sponsored website (22.37% vs 18.23%; *P*=.01).

Table 1. Smoking behavior and exposure to tobacco marketing among current and recent medical students (n=1659), Buenos Aires, Argentina, 2011.

Participant characteristics		n (%)
Age (years)		
	20-25	503 (30.32)
	26-27	618 (37.25)
	28-30	538 (32.43)
Gender		
	Male	447 (26.94)
	Female	1212 (73.06)
Student status		
	Currently enrolled student	924 (55.70)
	Graduated as physician	723 (43.58)
	Left medical school without degree	12 (0.72)
Smoking status		
	Never smoker	348 (20.98)
	Former smoker	858 (51.72)
	Current smoker	453 (27.31)
Depressive symptoms		
	Positive	870 (52.44)
Internet outcomes [a]		
	Has seen tobacco ad on Internet	446 (26.88)
	Received tobacco promotion[b]	482 (29.05)
	Used tobacco promotion[c]	167 (10.07)
	Accessed tobacco brand website	321 (19.35)

[a] “Do not remember” responses were categorized as “No”.

[b] Have received a promotion or invitation from a tobacco brand addressed specifically to the respondent.

[c] Have used a promotion or invitation from a tobacco brand addressed specifically to the respondent.

Smoking Behavior and Use of Tobacco Brand Websites

Most of the participants reported having smoked at least once (79.02%, 1311/1659) with 27.31% (453/1659) being current smokers. Table 2 presents the characteristics of current smokers and of those participants who ever accessed a tobacco-sponsored website. Among current smokers, 71.3% (323/453) were women and 58.3% (264/453) were current medical students. Most smokers (73.1%; 331/453) reported their first cigarette more than 60 minutes after waking up and 75.3% (341/453) smoked 10 or fewer cigarettes per day. A majority of smokers (73.1%, 331/453) intended to quit within 6 months and 32.9% (149/453) reported smoking in the medical school buildings. Almost half of them reported having received a tobacco promotion or invitation personally addressed to them (49.5%; 224/453) and 21.4% (97/453) used such promotion at least once. Among current smokers, 40.4% (183/453) reported accessing a tobacco brand website.

Almost 20% (321/1659) of the participants reported having accessed a tobacco brand website at least once in their lifetime. Of this group, 68.9% (221/321) were women, 60.4% (194/321) were current medical students, and 57.0% (183/321) were current smokers. Having seen a tobacco ad on the Internet was reported by 44.9% (144/321) of these participants; 68.5% (220/321) received a personally addressed tobacco marketing promotion or invitation while 39.9% (128/321) used one of the marketing promotions (Table 2). Almost all of them accessed these sites only when it was necessary for participating in a marketing promotion or using the promotion to obtain a gift (93.8%, 301/321). Most people logging on for marketing promotions reported entering once a month or less (58.9%, 189/321), while 25.5% (82/321) reported accessing the tobacco industry Internet sites once a week or more. Only 19 participants responded that they accessed the website for reasons other than marketing promotions. When the reason for going to the website was other than a marketing promotion, 73.7% (14/19) used the website once a month or less, while only 15.8% (3/19) used the sites once a week or more.

Table 2. Smoking behavior and use of tobacco brand websites among medical students and physicians (n=1659), Buenos Aires, Argentina, 2011.[a]

Participant characteristics		Current smoker, n (%) (n=453)	Accessed tobacco brand website, n (%) (n=321)
Age (years)			
	20-25	127 (28.0)	89 (27.7)
	26-27	175 (38.6)	113 (35.2)
	28-30	151 (33.3)	119 (37.1)
Gender			
	Male	130 (28.7)	100 (31.2)
	Female	323 (71.3)	221 (68.9)
Student status			
	Currently enrolled student	264 (58.3)	194 (60.4)
	Graduated as physician	184 (40.6)	123 (38.3)
	Left medical school without degree	5 (1.1)	4 (1.3)
Depressive symptoms			
	Positive	255 (56.3)	187 (58.3)
Internet outcomes [b]			
	Has seen tobacco ad on Internet	121 (26.7)	144 (44.9)
	Received tobacco marketing promotion[c]	224 (49.5)	220 (68.5)
	Used tobacco marketing promotion[d]	97 (21.4)	128 (39.9)

[a] Totals may not equal 100% due to rounding.

[b] “Do not remember” responses were categorized as “No”.

[c] Have received a marketing promotion from a tobacco brand addressed specifically to the respondent.

[d] Have used a marketing promotion from a tobacco brand addressed specifically to the respondent.

Multivariate Analysis

Unadjusted and adjusted logistic regression models assessed the association of age, gender, smoking status, student status, positive depression screen, daily use of Internet, seen tobacco ad on Internet, received a tobacco marketing promotion addressed to the participant, and used a tobacco marketing promotion addressed to the participant with the outcome of accessing a tobacco industry website. In these models, missing data for specific variables reduced the sample size for each model by no more than 2.7% (final n=1615). In adjusted models, participants were more likely to have accessed a tobacco brand website at least once if they were former or current smokers (OR 2.45, 95% CI 1.42-4.22 and OR 8.12, 95% CI 4.66-14.16 respectively), or if they reported having seen a tobacco advertisement on the Internet (OR 2.44, 95% CI 1.77-3.37), received a tobacco marketing promotion personally addressed to them (OR 5.62, 95% CI 4.19-7.55), or used one of these marketing promotions (OR 14.05, 95% CI 9.21-21.43) (Table 3).

Using similar adjusted models but with current smoking as the outcome, respondents were less likely to be current smokers if they reported having seen a tobacco advertisement on the Internet (OR 0.59, 95% CI 0.45-0.79) and more likely to be a smoker if they received a tobacco marketing promotion personally addressed to them (OR 2.64, 95% CI 2.02-3.45) or if they used one of these marketing promotions (OR 1.93, 95% CI 1.31-2.85).

Table 3. Predictors of medical students and physicians accessing tobacco brand websites, Buenos Aires, Argentina, 2011.

Variable		Referent	Access a tobacco brand website OR, 95% CI Unadjusted	Adjusted
Age[a]		Continuous	1.07 (1.001-1.14)[d]	1.10 (1.004-1.20)[d]
Smoking status [a]				
	Current smoker		13.29 (7.97-22.18)[e]	8.12 (4.66-14.16)[e]
	Former smoker	Never	2.97 (1.78-4.97)[e]	2.45 (1.42-4.22)[e]
Daily use of Internet [a]				
	Yes	No	1.43 (0.89-2.29)	1.14 (0.64-2.04)
Seen tobacco ad on Internet [a]				
	Yes	No	2.67 (2.07-3.45)[e]	2.44 (1.77-3.37)[e]
Received tobacco marketing promotion [b]				
	Yes	No	8.62 (6.57-11.31)[e]	5.62 (4.19-7.55)[e]
Used tobacco marketing promotion [c]				
	Yes	No	21.34 (14.46-31.50)[e]	14.05 (9.21-21.43)[e]

[a] These models are adjusted for the following variables: age, gender, smoking status, student status, depression, daily use of Internet, seen tobacco ad on Internet, received a tobacco marketing promotion addressed to them, and use a tobacco promotion addressed to them.

[b] This model is not adjusted for “used tobacco marketing promotion”.

[c] This model is not adjusted for “received tobacco marketing promotion”.

[d] $P<.05$

[e] $P \leq .001$

Discussion

Principal Findings

This is, to our knowledge, the first study that reports on the access and use of websites sponsored by the tobacco industry by medical students and physicians. Although other studies have described tobacco industry marketing promotions on the Web, we did not identify any that asked about the amount of use of these sites in a defined population. In the study population, nearly 20% of the respondents reported having accessed a website sponsored by the tobacco industry at least once, with almost all of them accessing the sites only when it was necessary for using or participating in a marketing promotion. This indicates that online marketing promotions for free gifts or tickets to cultural events are an effective way of delivering tobacco advertising and reaching a young adult population—even medical students and recently graduated physicians. This study also found that participants who received a marketing promotion or an invitation personally addressed to them were more likely to be current smokers and to have accessed one of these websites, even when adjusting for other variables. Intriguingly, respondents who reported having seen a tobacco advertisement on the Internet were less likely to be current smokers. One possible explanation for this observation would be that people who do not smoke are more aware of tobacco advertising than those who do smoke and thus more likely to report seeing it on a survey.

The implications of these findings are that the tobacco industry identifies ways to market to their target population while voluntarily accepting “advertising bans” in print media, radio, and television. Tobacco industry marketing has been regulated over the past decades in most countries, driven in part by the WHO Framework Convention on Tobacco Control endorsing “a comprehensive ban on advertising, marketing promotion, and sponsorship” as a way to reduce tobacco consumption. In fact, implementing a complete advertising ban is one of the evidence-based policies included in modeling full implementation of tobacco control policies in Argentina [30]. However, the Internet remains mostly unregulated for tobacco advertising, marketing promotion, and sponsorship and thus undermines the purpose of those restraints. Our study suggests that tobacco industry online strategies let companies interact directly with consumers, allowing for active user participation, and avoid the advertisement bans being proposed and implemented.

Health professionals possess the greatest potential of any occupation to promote a reduction in tobacco use at the population level [31]. Physicians and other health professionals may contribute to tobacco control by acting as role models, by providing counseling and smoking cessation treatment, and by publically advocating for comprehensive tobacco control public policies [32]. Moreover, physician smoking status may affect their willingness to initiate cessation interventions in their patients and their effectiveness when providing counseling [12-14]. The results of this study indicate that the prevalence of current smoking among medical students and physicians was lower than the one found in the 2005 medical student survey (27.3% vs 35.5%) [17]. This seems to be consistent with the trend toward lower smoking rates in Argentina's general population, which has shown a reduction over four years based on the national risk factor surveys [2]. It also gets closer to the prevalence found in other South American countries [21]. However, although our sample denominator was large, it was not limited to third-year students only.

Limitations

Our study has several limitations to consider. First was the relatively low response rate of the sample denominator to conduct the survey. Although this collaboration rate is similar to other studies using Web-based surveys [20,33,34], generalizations based on these data need to be cautious. Unfortunately, we do not have information about the demographics of non-respondents in order to compare their characteristics with the participants who actually completed the survey. However, the study population was mostly women and this reflects the composition of the UBA medical school where women account for two-thirds of the student body [22]. The proportion by age groups is also what would be expected based on UBA enrollment. Another limitation is the fact that we did not ask how many times a person accessed one of these websites (only once or more) and when they actually accessed the sites. Thus, we cannot be certain that the students were current smokers at the time of website access. Finally, use of cross-sectional data prevents establishing any causal inference or temporal relationships from our results.

Conclusions

Despite these limitations, our study suggests that tobacco industry websites do reach young adult medical students and physicians in a middle-income country with their marketing promotions. This strategy may be essential in maintaining current smoking status in this group and young adults in general. Internet advertisement may also give the tobacco industry an opportunity to improve their public image while undermining the dissemination of accurate public health information. In June 2011, Argentina's Congress approved a law that regulates tobacco products [35]. The text of the law bans publicity, marketing promotions, and sponsorship of any tobacco product, direct or indirect, through any media. Although the law has now been implemented by the Executive Branch, our study indicates the need for further research to analyze the impact of this law on tobacco-sponsored websites and social media communication. Legislation to ban tobacco advertising and marketing promotions needs to explicitly include Internet sites and related social media.

Acknowledgments

This research was funded by grant No.TW05935 from the Tobacco Research Network Program, Fogarty International Center, National Cancer Institute, National Institute of Drug Abuse, and National Institutes of Health, and by grant No. 001726-037 from Research on International Tobacco Control, International Development Research Center, Canada.

We thank Professor Rodolfo Rothlin from the University of Buenos Aires for assistance during data collection, Cecilia Populus-Eudave for administrative and research support, and Steven E Gregorich, PhD, for advice on analytic approaches at the University of California, San Francisco.

Authors' Contributions

MVS was the lead author in conceiving the project and development of the survey with supervision and mentoring by RM, CK, and EPS. MVS implemented the survey, cleaned the data, and conducted all data analysis. All authors contributed to writing of the manuscript.

Conflicts of Interest

None declared.

References

1. World Health Organization. Fact Sheet: Tobacco URL: http://www.who.int/mediacentre/factsheets/fs339/en/index.html [accessed 2013-12-18] [WebCite Cache ID 6CxzzuLD9]
2. Ferrante D, Linetzky B, Konfino J, King A, Virgolini M, Laspiur S. 2009 National Risk Factors Survey: Evolution of the epidemic of chronic non-communicable diseases in Argentina. Cross-sectional study. Revista Argentina de Salud Pública 2011;2(6):34-41.
3. Mejia R, Pérez-Stable EJ. Tobacco epidemic in Argentina: The cutting edge of Latin America. Prevention and Control 2006 Jan;2(1):49-55. [doi: 10.1016/j.precon.2006.04.003]
4. Freeman B, Chapman S. Gone viral? Heard the buzz? A guide for public health practitioners and researchers on how Web 2.0 can subvert advertising restrictions and spread health information. J Epidemiol Community Health 2008 Sep;62(9):778-782. [doi: 10.1136/jech.2008.073759] [Medline: 18701726]
5. Freeman B, Chapman S. British American Tobacco on Facebook: Undermining Article 13 of the global World Health Organization Framework Convention on Tobacco Control. Tob Control 2010 Jun;19(3):e1-e9 [FREE Full text] [doi: 10.1136/tc.2009.032847] [Medline: 20395406]

6. Elkin L, Thomson G, Wilson N. Connecting world youth with tobacco brands: YouTube and the Internet policy vacuum on Web 2.0. Tob Control 2010 Oct;19(5):361-366. [doi: 10.1136/tc.2010.035949] [Medline: 20739706]
7. Forsyth SR, Malone RE. "I'll be your cigarette--light me up and get on with it": examining smoking imagery on YouTube. Nicotine Tob Res 2010 Aug;12(8):810-816 [FREE Full text] [doi: 10.1093/ntr/ntq101] [Medline: 20634267]
8. Instituto Nacional de Estadistica y Censos. Buenos Aires, Argentina Accesos a internet: primer trimestre 2011 URL: http://www.indec.gov.ar/nuevaweb/cuadros/14/internet_06_11.pdf [accessed 2013-12-18] [WebCite Cache ID 6Cxzs8BzX]
9. World Bank. Internet users (per 100 people): 2012 URL: http://data.worldbank.org/indicator/IT.NET.USER.P2 [accessed 2013-12-18] [WebCite Cache ID 6LxkQbo0F]
10. Freeman B, Chapman S. Open source marketing: Camel cigarette brand marketing in the "Web 2.0" world. Tob Control 2009 Jun;18(3):212-217. [doi: 10.1136/tc.2008.027375] [Medline: 19211615]
11. World Health Organization: Tobacco Free Initiative. Code of practice on tobacco control for health professional organizations URL: http://www.who.int/tobacco/wntd/2005/codeofpractice/en/index.html [accessed 2013-12-18] [WebCite Cache ID 6Cy09juwk]
12. Meshefedjian GA, Gervais A, Tremblay M, Villeneuve D, O'Loughlin J. Physician smoking status may influence cessation counseling practices. Can J Public Health 2010;101(4):290-293. [Medline: 21033533]
13. Pipe A, Sorensen M, Reid R. Physician smoking status, attitudes toward smoking, and cessation advice to patients: an international survey. Patient Educ Couns 2009 Jan;74(1):118-123. [doi: 10.1016/j.pec.2008.07.042] [Medline: 18774670]
14. Ulbricht S, Baumeister SE, Meyer C, Schmidt CO, Schumann A, Rumpf HJ, et al. Does the smoking status of general practitioners affect the efficacy of smoking cessation counselling? Patient Educ Couns 2009 Jan;74(1):23-28. [doi: 10.1016/j.pec.2008.07.047] [Medline: 18818045]
15. Tong EK, Strouse R, Hall J, Kovac M, Schroeder SA. National survey of US health professionals' smoking prevalence, cessation practices, and beliefs. Nicotine Tob Res 2010 Jul;12(7):724-733. [doi: 10.1093/ntr/ntq071] [Medline: 20507899]
16. Zylbersztejn H, Cardone A, Vainstein N, Mulassi A, Calderón J, Blanco P, et al. Tabaquismo en médicos de la República Argentina: Estudio TAMARA. Revista Argentina de Cardiologia 2007;75(2):109-116.
17. GTSS Collaborative Group. Tobacco use and cessation counselling: Global health professionals survey pilot study, 10 countries, 2005. Tob Control 2006 Jun;15 Suppl 2:ii31-ii34 [FREE Full text] [doi: 10.1136/tc.2006.015701] [Medline: 16731522]
18. Lewis KE, Shin D, Davies G. Smoking habits and attitudes toward tobacco bans among United Kingdom hospital staff and students. Int J Tuberc Lung Dis 2011 Aug;15(8):1122-1126. [doi: 10.5588/ijtld.10.0783] [Medline: 21740679]
19. Kusma B, Quarcoo D, Vitzthum K, Welte T, Mache S, Meyer-Falcke A, et al. Berlin's medical students' smoking habits, knowledge about smoking and attitudes toward smoking cessation counseling. J Occup Med Toxicol 2010;5:9 [FREE Full text] [doi: 10.1186/1745-6673-5-9] [Medline: 20398350]
20. Morrell HE, Cohen LM, Dempsey JP. Smoking prevalence and awareness among undergraduate and health care students. Am J Addict 2008;17(3):181-186 [FREE Full text] [doi: 10.1080/10550490802019899] [Medline: 18463994]
21. Warren CW, Sinha DN, Lee J, Lea V, Jones NR. Tobacco use, exposure to secondhand smoke, and cessation counseling among medical students: cross-country data from the Global Health Professions Student Survey (GHPSS), 2005-2008. BMC Public Health 2011;11:72 [FREE Full text] [doi: 10.1186/1471-2458-11-72] [Medline: 21284864]
22. Secretaría de Asuntos Académicos. Universidad de Buenos Aires. Censo de Estudiantes 2004 URL: http://www.uba.ar/institucional/censos/Estudiantes2004/censo-estudiantes.pdf [accessed 2013-12-18] [WebCite Cache ID 6Cy0OvNUf]
23. GTSS Collaborative Group. The global tobacco surveillance system. Tob Control 2006 Jun;15 Suppl 2:ii1-ii3 [FREE Full text] [doi: 10.1136/tc.2006.015719] [Medline: 16731520]
24. Alderete E, Kaplan CP, Gregorich SE, Mejía R, Pérez-Stable EJ. Smoking behavior and ethnicity in Jujuy, Argentina: evidence from a low-income youth sample. Subst Use Misuse 2009;44(5):632-646 [FREE Full text] [doi: 10.1080/10826080902809717] [Medline: 19360537]
25. Primack BA, Gold MA, Switzer GE, Hobbs R, Land SR, Fine MJ. Development and validation of a smoking media literacy scale for adolescents. Arch Pediatr Adolesc Med 2006 Apr;160(4):369-374 [FREE Full text] [doi: 10.1001/archpedi.160.4.369] [Medline: 16585481]
26. Salgado MV, Pérez-Stable EJ, Primack BA, Kaplan CP, Mejia RM, Gregorich SE, et al. Association of media literacy with cigarette smoking among youth in Jujuy, Argentina. Nicotine Tob Res 2012 May;14(5):516-521 [FREE Full text] [doi: 10.1093/ntr/ntr240] [Medline: 22193569]
27. Murphy JM, Horton NJ, Monson RR, Laird NM, Sobol AM, Leighton AH. Cigarette smoking in relation to depression: historical trends from the Stirling County Study. Am J Psychiatry 2003 Sep;160(9):1663-1669. [Medline: 12944343]
28. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
29. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. J Gen Intern Med 1997 Jul;12(7):439-445 [FREE Full text] [Medline: 9229283]
30. Konfino J, Ferrante D, Mejia R, Coxson P, Moran A, Goldman L, et al. Impact on cardiovascular disease events of the implementation of Argentina's national tobacco control law. Tob Control 2012 Oct 29:1. [doi: 10.1136/tobaccocontrol-2012-050599] [Medline: 23092886]

31. World Health Organization: Tobacco Free Initiative. Health systems and professionals URL: http://www.who.int/tobacco/control/health_professionals/en/index1.html [accessed 2013-12-18] [WebCite Cache ID 6Cy0p934V]
32. American Cancer Society. 2003. Engaging doctors in tobacco control URL: http://www.paho.org/english/ad/sde/ra/Guide2a_Engaging_Doctors_in_Tobacco_Control.pdf [accessed 2013-12-18] [WebCite Cache ID 6Cy0x4u1n]
33. Hayslett MM, Wildemuth BM. Pixels or pencils? The relative effectiveness of Web-based versus paper surveys. Library & Information Science Research 2004;26(1):73-93. [doi: 10.1016/j.lisr.2003.1011.1005]
34. Sinclair M, O'Toole J, Malawaraarachchi M, Leder K. Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. BMC Med Res Methodol 2012;12:132 [FREE Full text] [doi: 10.1186/1471-2288-12-132] [Medline: 22938205]
35. InfoLEG. 2011. Tabaco: Ley Nacional 26.687 URL: http://infoleg.mecon.gov.ar/infolegInternet/anexos/180000-184999/183207/norma.htm [accessed 2013-12-18] [WebCite Cache ID 6Cy1COXOH]

Abbreviations

GHPS: Global Health Professionals Survey
UBA: University of Buenos Aires
WHO: World Health Organization

Edited by G Eysenbach; submitted 15.01.13; peer-reviewed by J Brown, K Wang; comments to author 08.04.13; revised version received 28.05.13; accepted 30.10.13; published 06.02.14.

Please cite as:
Salgado MV, Mejia R, Kaplan CP, Perez-Stable EJ
Smoking Behavior and Use of Tobacco Industry Sponsored Websites Among Medical Students and Young Physicians in Argentina
J Med Internet Res 2014;16(2):e35
URL: http://www.jmir.org/2014/2/e35/
doi:10.2196/jmir.2528
PMID:24509433

Review

A Decade of Research on the Use of Three-Dimensional Virtual Worlds in Health Care: A Systematic Literature Review

Reza Ghanbarzadeh[1]; Amir Hossein Ghapanchi[1], PhD; Michael Blumenstein[1], PhD; Amir Talaei-Khoei[2,3], PhD

[1]School of Information and Communication Technology, Griffith University, Gold Coast, Queensland, Australia
[2]Agile Information Systems Lab, School of Systems, Management and Leadership, University of Technology Sydney, Sydney, Australia
[3]School of Business, Faculty of Arts and Business, University of the Sunshine Coast, Maroochydore DC, Australia

Corresponding Author:
Reza Ghanbarzadeh
School of Information and Communication Technology
Griffith University
G23 2.32, Gold Coast Campus
Gold Coast, Queensland, 4222
Australia
Phone: 61 7 555 28502
Fax: 61 7 555 28502
Email: reza.ghanbarzadeh@griffithuni.edu.au

Abstract

Background: A three-dimensional virtual world (3DVW) is a computer-simulated electronic 3D virtual environment that users can explore, inhabit, communicate, and interact with via avatars, which are graphical representations of the users. Since the early 2000s, 3DVWs have emerged as a technology that has much to offer the health care sector.

Objective: The purpose of this study was to characterize different application areas of various 3DVWs in health and medical context and categorize them into meaningful categories.

Methods: This study employs a systematic literature review on the application areas of 3DVWs in health care. Our search resulted in 62 papers from five top-ranking scientific databases published from 1990 to 2013 that describe the use of 3DVWs for health care specific purposes. We noted a growth in the number of academic studies on the topic since 2006.

Results: We found a wide range of application areas for 3DVWs in health care and classified them into the following six categories: academic education, professional education, treatment, evaluation, lifestyle, and modeling. The education category, including professional and academic education, contains the largest number of papers (n=34), of which 23 are related to the academic education category and 11 to the professional education category. Nine papers are allocated to treatment category, and 8 papers have contents related to evaluation. In 4 of the papers, the authors used 3DVWs for modeling, and 3 papers targeted lifestyle purposes. The results indicate that most of the research to date has focused on education in health care. We also found that most studies were undertaken in just two countries, the United States and the United Kingdom.

Conclusions: 3D virtual worlds present several innovative ways to carry out a wide variety of health-related activities. The big picture of application areas of 3DVWs presented in this review could be of value and offer insights to both the health care community and researchers.

(J Med Internet Res 2014;16(2):e47) doi:10.2196/jmir.3097

KEYWORDS
3D virtual worlds; 3DVW; health care

Introduction

Since the introduction of Web 2.0, there has been a significant improvement in the potential of Web applications. Web 2.0, as a new technology, facilitates activities such as collaboration, interaction, social networking, and participation between users [1,2]. One of the important applications of Web 2.0 is the three-dimensional virtual world (3DVW). A 3DVW is a computer-generated, simulated, networked, graphic, and multimedia environment, usually running on the Web and designed so that users can “live in” and interact via their own digital and graphical self-representations known as “avatars”

[3,4]. In 3DVWs, the avatars of multiple users can interact with one another through text or voice tools, either privately or publicly, inside the various regions. By using “serious gaming” formats, avatars can move through the environments and worlds and manipulate objects. They share most of the capabilities of virtual reality (VR) technologies, especially the rendering of 3D cyberspace, they are more accessible to users via Internet-connected high-speed personal computers, and as a multi-user environment, they support social networking and interactivity.

Since the generation of 3DVWs, their lands have grown significantly with millions of residents joining globally for a variety of purposes. They offer platforms for business, education, communication, and organizational developments. As a tool for innovation, 3DVWs attract not only individuals, but also organizations, universities, companies, corporations, government agencies, and private groups that take advantage of the opportunities for collaboration, education, and communication. They are the next evolution of the Internet and social media, and they merge most of the qualities of the Web, telecommunication technology, online gaming, distance learning, social networking, creative applications, and user-generated contents.

For example, according to Linden Lab (2013), over the last 10 years, 36 million accounts have been created in Second Life (a 3DVW where users can socialize, connect, and create using voice and text chat) [5] and US $3.6 billion spent on virtual possessions. The total time users have so far spent on Second Life is equivalent to almost 217,266 years. Currently, more than a million people visit Second Life every month, and approximately 400,000 new registrations are created monthly. On average, 1.2 million daily transactions are conducted for virtual possessions, 2.1 million user-created virtual possessions are for sale, and Second Life’s landmass is nearly 700 square miles [6].

For online users, 3DVWs provide opportunities to explore, create, imagine, collaborate, role play, interact, socialize, learn, and experience events in a safe and vivid manner, and they can also be linked to the real world and other Web resources and services in a variety of scenarios. They offer improved experiences to users in relation to health care information and education, skill-building, group support, and individual consultation in terms of health. Over the last decade, there has been growing interest among the medical and public health communities in using 3DVWs for treatment, education, development, and simulation. There is a vast variety of medical and health-related 3D virtual environments inside these worlds. They currently feature a number of medical and health-related projects, have been a key area of growth, and may offer opportunities for patients, physicians, providers, educators, and health care institutions to improve both the quality and efficiency of care, treatment, and education.

Several studies have been conducted on this topic. Kamel Boulos et al introduced 3DVWs and their educational potential to medical and health educators and librarians [7]. Hansen provided an overview of 3DVWs currently used in health care professional education and medicine [8]. A survey of health-related activities on Second Life has been provided by Peck and Miller [9]. The opportunities available to nursing students within a multi-user virtual environment are presented by Beard et al [10].

The health care industry has developed tremendously due to the growing implementation of health information technology infrastructures. 3DVWs, as one of these technologies, offer great opportunities to the health community, and the health care sector are currently starting to migrate to these kinds of platforms as this technology is gradually maturing and becoming more popular and affordable. Health care organizations, universities, groups, and individuals are currently using 3DVWs for a range of clinical and health-related activities and purposes. However, the benefits of 3DVWs in health and medicine are less well understood to the health care community and researchers. It is important for them to understand the impact of 3DVWs on their field so they can consider the advantages of this technology in their own research, business, or profession. The significance of this technology in health, the remarkable attention from academia, and the lack of a comprehensive systematic literature review in this field motivated us to survey recent literature and attempt to create a big picture of 3DVWs and their application in health care. This paper attempts to answer the following three research questions:

1. What fields of health care research have been targeted by applications of 3DVW technologies?
2. In each of these fields, in what contexts have 3DVW technologies been applied?
3. What are the contributions of 3DVW technologies in each of these contexts?

Methods

Summary

To study the efforts made in the application of 3DVWs in the health care sector, we performed a systematic literature review, which is a methodical approach to the identification, evaluation, and interpretation of previous studies conducted on a specific research topic [11]. Our study follows the Kitchenham [11] guidelines for performing a systematic literature review. In this section, we describe the steps in the research methodology and identify the process of inclusion and exclusion of papers, as well as data extraction and analysis during the search process.

Search and Selection Procedures

The search process for relevant literature was completed in six stages. Figure 1 indicates the stages of study selection performed in accordance with Kitchenham [11] and Ghapanchi [12] guidelines.

In the first stage, we identified five main scientific databases in which to search for our keywords. In Stage 2, we performed a search of these five databases for 42 keywords. As of July 20, 2013, our database searches revealed 1088 primary studies. As shown in Figure 1, Stages 3 to 5 were performed twice in our search process. In the first iteration, Stage 3, we excluded 789 papers on the basis of their titles. In Stage 4, on the basis of abstracts (in some cases we had to read the full text to identify unrelated papers), 262 papers were excluded from the remaining

299 papers. The total number of remaining relevant papers was 37; however, this number of papers was not sufficient to generate an adequate conclusion. In order to increase the comprehensiveness of our study, we needed to find more papers related to this topic. According to Kitchenham [11] guidelines, one way to achieve this goal is to investigate the references of those 37 papers. Kitchenham suggests that this process can be repeated until achieving a reasonable number of papers. Therefore, we investigated the referenced papers of these 37 papers in Stage 6. This yielded an additional 1183 papers. Some 1152 papers were discarded on the basis of their titles. Another six papers were excluded because of their abstracts, resulting in 25 remaining papers. Altogether, as of August 20, 2013, we had 2271 papers, of which 2209 were discarded as being unrelated to our topic. Finally, at the end of phase two, the total number of papers was 62, which seemed reasonable for generating an appropriate conclusion. We therefore stopped the search process at phase two. Other authors have executed similar processes [12,13]. No paper was excluded in either iteration on the basis of the full text. Table 1 represents a summary of our search and the exclusion process in each stage.

Figure 1. Stages in the study selection procedure.

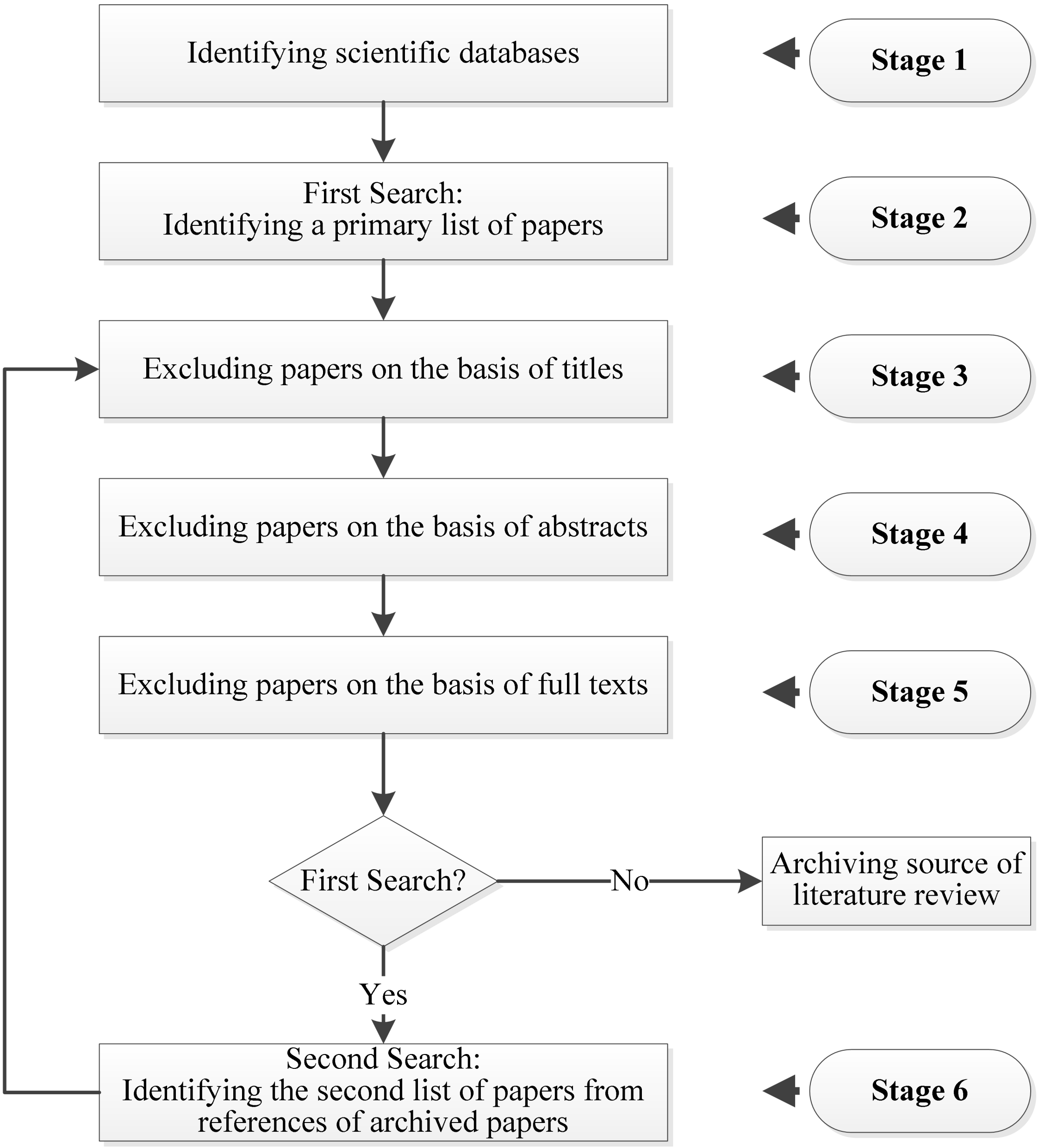


Table 1. Summary of search and paper selection process.

	First search	References included
Initial number of papers	1088	1183
Excluded by title	789	1152
Excluded by abstract	262	6
Final number of remaining papers	37	25
Total number of remaining papers	62	

Resources Searched

The following scientific databases were used in the search for the keywords noted in Multimedia Appendix 1: ScienceDirect, ProQuest, IEEE Explore, ACM Digital Library, and PubMed.

Table 2 shows the primary search results for the five scientific databases for searching in “all fields”, “title”, and “title/abstract/keyword”, separately.

Table 2. Summary of initial search results for each database.

Database	Searched in	Number of papers found
Science Direct	Title/Abstract/Key	175
	Title	3
	All fields	13998
ProQuest	Everywhere except full text	116
	Title	0
	All fields	3119
PubMed	Title/Abstract	479
	Title	15
	All fields	760
IEEE Xplore	Abstract	261
	Title	7
	All fields	745
ACM Digital Library	Abstract	57
	Title	2
	All fields	6962

Search Terms

The advanced search service provided by each scientific database search engine was used to perform the search operations. According to the search patterns offered by each search engine, the title, abstract, keywords, and, in some cases, the full text of articles were sourced by means of 42 search terms. Multimedia Appendix 1 shows the search terms used in each of the scientific databases. During all search processes, a publication date filtration was carried out either in the search queries or in the advanced search options.

Inclusion and Exclusion Criteria

In order to select materials for this systematic review, some inclusion and exclusion criteria were considered. It was important for us to investigate different perspectives of applications of 3DVWs in the health care sector; therefore, we reviewed both of the design guidelines and empirical evidences for applicability in this systematic review. Accordingly, both of the empirical and technical studies have been targeted by this search. We included studies published between January 1990 and July 2013 and excluded studies in languages other than English. According to the search results, there were no relevant papers between the years 1990 and 2005.

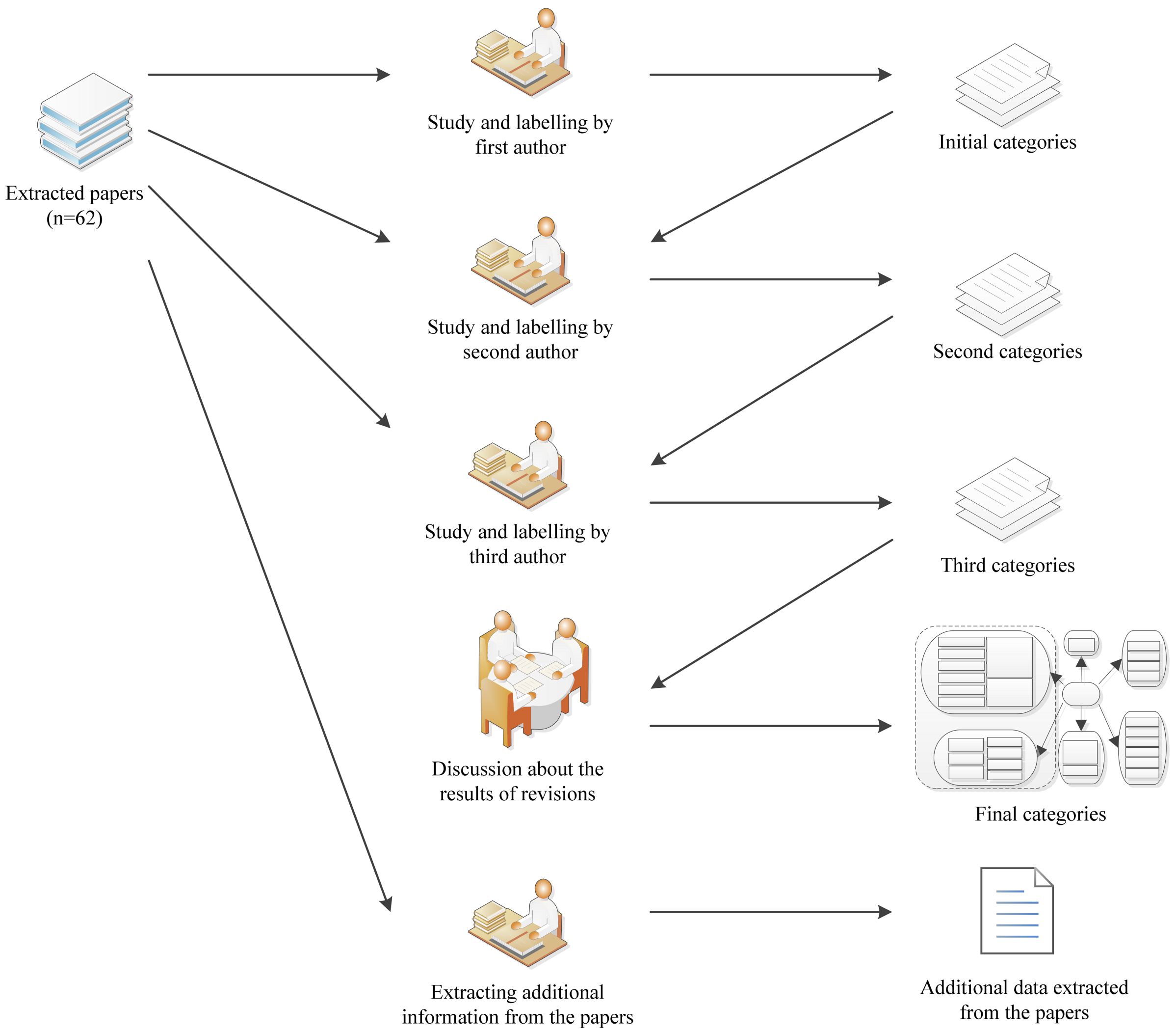
Data Extraction

Two types of data were extracted from 62 studies in this systematic review: (1) 3DVWs and their applications in health-related activities, and (2) year of publication, country, affiliated department, and names of sources. Consistent with health care and medical purposes, 3DVWs were designed, implemented, and applied in a majority of main categories, contexts, and subcontexts of application.

Data Analysis

Our purpose in this study was to group whole papers into meaningful categories of applications in health care. To perform a data analysis, we read the title, abstract, and full text of the 62 extracted papers and tried to classify each paper in an appropriate category based on focus, main area of research, and context. To accomplish this, we performed three reviews. First, each paper was reviewed completely by one of the authors and an appropriate label allocated according to the application area of 3DVWs. At the end of the first stage, all of the papers were classified in different categories. To clarify the classification, the process was repeated for the second and third times by other authors. Afterwards, the results of the first, second, and third revisions were discussed in a meeting of 5 people including the authors. During the meeting, labels were revised and some of the categories were merged. Finally, we grouped all of the extracted papers into six major research categories based on the application areas of 3DWVs. Some of the papers could have been placed in either of two different categories so we had to select the more appropriate one. After this, we attempted to extract additional information from the articles, such as contexts, subcontexts, year of publication, and country of publication. Figure 2 demonstrates the data analysis process.

Figure 2. Stages in data analysis procedure.



Results

Summary

Figure 3 illustrates the distribution of published papers per year. Between 1990 and 2005, there were no papers covering the application of 3DVWs in health care, primarily because 3DVWs are new technologies and the early editions of these virtual worlds were created and released in 2003. The number of published papers per year rose markedly from 1 to 13 between 2006 and 2010, decreasing to nine in 2009. There was a decline to ten papers in 2011 and eight in 2012. Our research took place in July 2013, and three papers had been published this year prior to this date.

Figure 4 shows the percentages of the extracted papers from different countries around the world. Around 60% of the total papers (37/62) were from the United States. The United Kingdom is the next source of papers in this field with 13% of publications (8/62). In third place, at 5% each (3/62), are Australia, Italy, and Canada. The remaining 12% of papers (8/62) come from Sweden, Hong Kong, Korea, Japan, Israel, and New Zealand.

Although Figure 4 does not reflect research capability, it compares papers geographically. Its main purpose is to provide insight into the amount of research done in different countries. We believe that most of the research in this field has been done in North America and may not be applicable in different health care settings, social constructs, cultural contexts, etc. Therefore, we recommend comparative research in different geographic locations.

In order to perform a systematic review of 3DVW-related studies, we classified by field of application areas of 3DVW in health care. To provide this classification, we read all papers in our study and then attempted to assign an appropriate label to each one. In the second and third revisions, we investigated these labels and made some modifications. Finally, we found six main topics in the literature: (1) treatment, (2) modeling, (3) evaluation, (4) lifestyle, (5) academic education, and (6) professional education (see Figure 5; we grouped academic education and professional education categories using a dashed line). Academic and professional education could have been placed in a broad category of education. But to make a more accurate and detailed classification, we broke this large category to two smaller categories according to their context. The six main categories as well as their associated contexts and subcontexts are described in the following sections.

Figure 3. Frequency of papers per year.

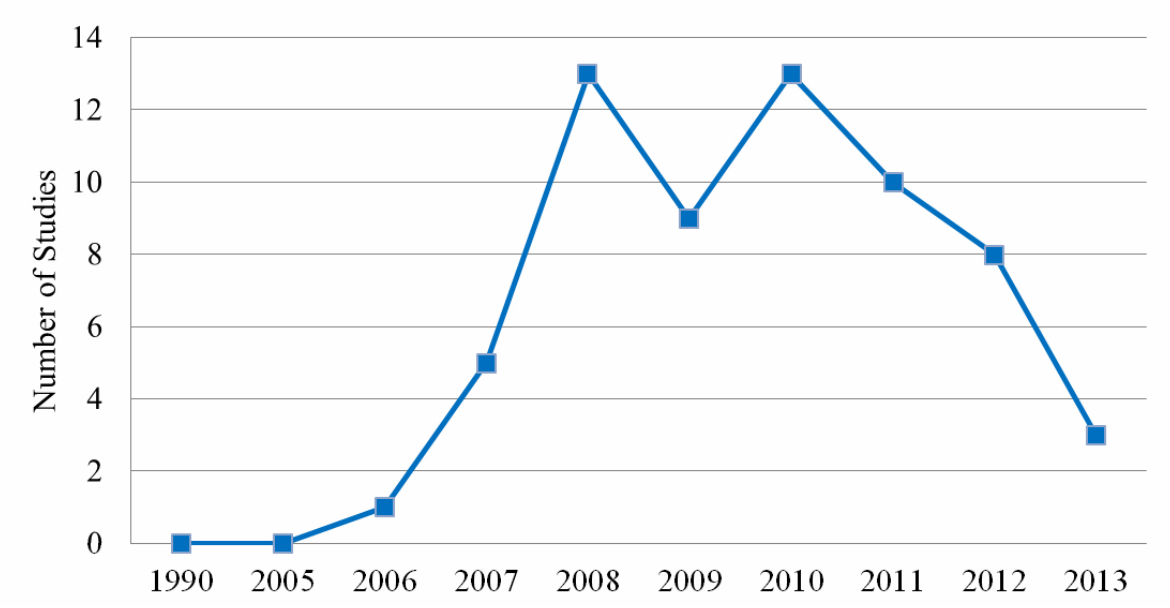


Figure 4. Frequency of papers from different countries.

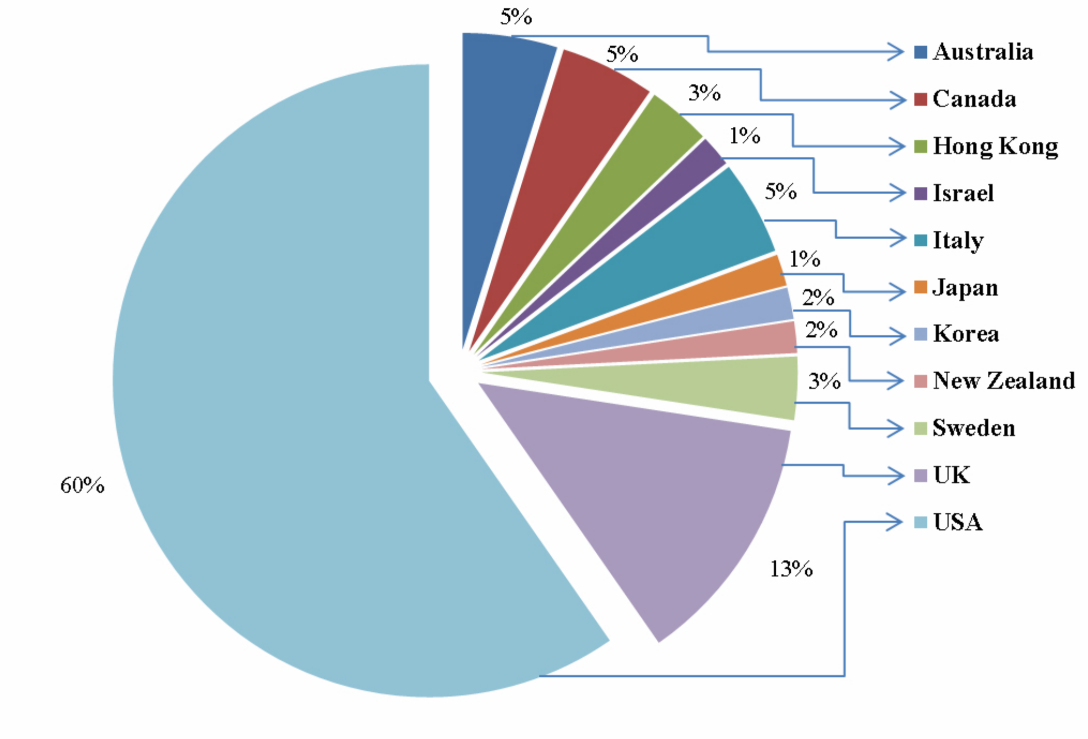
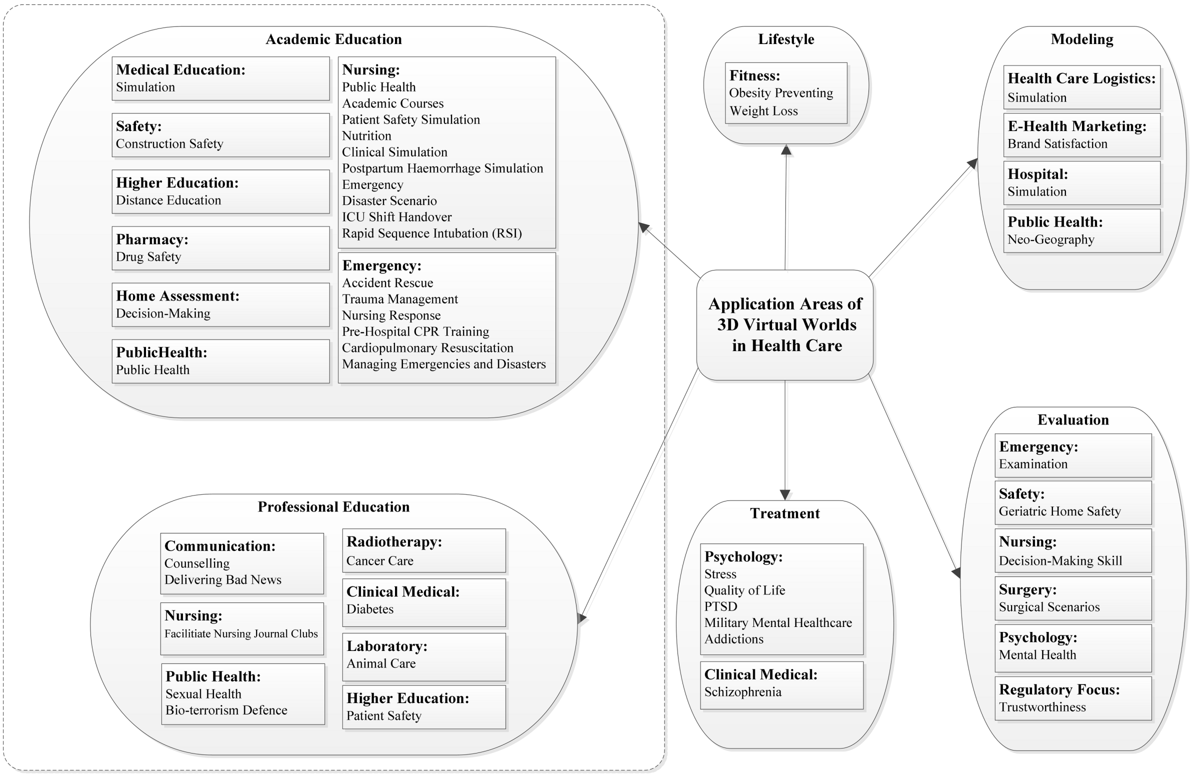


Figure 5. Taxonomy of application areas of 3DVWs in health care.



Treatment

Studies in this category investigated different applications of 3DVWs for treatment and therapy purposes in health care. Nine studies in this category addressed the different use of these worlds in various health care treatment applications. In most cases, patients, nurses, physicians, or other medical staff had their own avatars in a specific environment in a VW, and patients were treated with specialized techniques. Treatment was performed in two health care contexts: psychological treatment and clinical treatment. For example, Gorini et al addressed the use of 3DVWs in online mental health applications and addiction treatment [14]. A method for implementing virtual environments to train people with schizophrenia to develop conversational skills in specific situations was reported by Ku et al [15]. Linton et al used Second Life to reduce stress, anxiety, and psychological distress [16].

Table 3 shows the various contexts and subcontexts related to studies for which 3DVWs were used for treatment purposes [14-22]. Most of the treatment programs were carried out on psychological and mental patients.

In order to provide meaningful insights into each application, Table 4 gives brief descriptions of each study. For example, a new approach based on using 3DVWs was described by Pioggia et al for the assessment and treatment of psychological stress [17]. A drug addiction treatment program using Second Life was introduced by Gustafson et al [22].

Table 3. Health care contexts and subcontexts in the treatment category.

Context	Subcontext/Reference
Psychology	Psychological stress [17]
	Quality of life of cancer patients [18]
	PTSD [19]
	Military mental health care [20]
	Stress reduction program [16]
	PTSD [21]
	Different addictions [14]
	Drug addiction [22]
Clinical/medical	Schizophrenia [15]

Table 4. Brief descriptions of papers in the treatment category.

Subcontext	Brief description/quote
Psychological stress	The authors describe a new approach to assessment and treatment of psychological stress based on 3DVW [17].
QoL of cancer patients	“This research investigated the effect of using a three-dimensional online chat environment on community development for cancer patients’ mental health” p. 490 [18].
PTSD	“[…a new Virtual Reality Software Suite, ARGAMAN is introduced which] is an innovative, unique solution providing an immersive virtual reality therapeutic system for treatment of PTSD (Post-Traumatic Stress Disorder) in people who witnessed or suffered terror attacks or other traumatic experiences” p. 34 [19].
Military mental health care	“…presents the design vision for a Clinical VR project called SimCoach that aims to create intelligent virtual human agents to serve the role of online health care guides/coaches for military Service Members, Veterans and their significant others in an effort to break down barriers to care” p. 504 [20].
Stress reduction program	“In this pilot study, [authors explored] the feasibility of translating a face-to-face stress reduction program into an online virtual setting and estimate the effect size of the intervention” p. 1 [16].
PTSD	“In this article, [authors] suggest that the use of a new technological paradigm, Interreality, may improve the clinical outcome of PSTV [Posttraumatic Stress Disorders]” p. 55 [21].
Different addictions	“This paper describes the development and implementation of a form of tailored immersive e-therapy called p-health whose key factor is interreality, that is, the creation of a hybrid augmented experience merging physical and virtual worlds” p. 1 [14].
Drug addiction	A drug addiction treatment program using Second Life [22]
Schizophrenia	“In this paper, [authors] report on a method of implementing virtual environments (VEs) in order to train people with schizophrenia to develop conversational skills in specific situations, which could overcome the shortcomings of or complement conventional role-playing techniques” p. 567 [15].

Modeling

Modeling is an activity that aims to help one understand, define, quantify, visualize, or simulate a specific part or feature in the real world in a simpler way. Four studies in our systematic review examined the application of 3DVWs for modeling purposes in health care. They applied different 3DWVs for modeling of hospitals, health care logistics, eHealth marketing, and public health. In most cases, the virtual replica of a medical environment such as a university lab, hospital, hospital ward, emergency ward, or operating room were modeled and designed inside the various virtual islands using virtual worlds. The main purpose of modeling is the simulation of real-world environments for different purposes, for example, training, assessment, and examining. Kamel Boulos et al implemented a complete hospital using Second Life for several health care purposes [23]. A virtual smart hospital with tools, instruments, virtual RFID (radio-frequency identification) check, and real-time location-aware system for tracking all items was implemented by Thompson and Hagstrom using Second Life [24]. Table 5 indicates the health contexts and subcontexts of papers using 3DVWs for modeling goals [23-26].

Table 6 gives brief descriptions for each paper from the extracted studies in this category. For instance, using 3DVWs, Jin and Lee examined the effects of the regulatory fit that consumers experience in interactive eHealth marketing on their brand satisfaction and brand trust [25]. Kamel Boulos et al used Google Maps in Second Life for public health neo-geography purposes [26].

Table 5. Health care contexts and subcontexts in the modeling category.

Context	Subcontext/Reference
Health care logistics	Simulation [24]
eHealth marketing	Brand satisfaction and brand trust [25]
Hospital	Simulation [23]
Public health	Neo-geography [26]

Table 6. Brief descriptions of papers in the modeling category.

Subcontext	Brief description/quote
Simulation	A virtual smart hospital with tools, instruments, virtual RFID check, and real-time location-aware system for tracking all items [24].
Brand satisfaction and brand trust	“This study examined the effects of the regulatory fit that consumers experience in interactive e-health marketing on their brand satisfaction and brand trust” p. 673 [25].
Simulation	“This paper provides a tour of several flagship Web 3D experiences in Second Life, including Play2Train Islands (emergency preparedness training), the US Centers for Disease Control and Prevention- CDC Island (public health), Karuna Island (AIDS support and information), Tox Town at Virtual NLM Island (US National Library of Medicine - environmental health), and Jefferson’s Occupational Therapy Center” p. 290 [23].
Neo-geography	Demonstrates the use of Google maps in Second Life for public health neo-geography [26].

Evaluation

Studies in this category focus on assessment and evaluation by 3DVWs in health care. In most cases, these worlds were used for evaluation of a particular proficiency in specific groups such as nurses and supervisors, assessment of safety in specific environments, the measurement of factors in emergency services, or investigating the rate of improvement in a patient. The main health contexts for which 3DVWs have been used for evaluation are nursing, surgery, psychology, safety regulatory focus, and emergency.

For example, Patel et al used Second Life for performance assessment of 63 surgeons in clinical scenarios [27]. In 2012, 3DVWs were used by Andrade et al to evaluate the feasibility, usability, and acceptability of virtual worlds for geriatric home safety assessments [28]. Another instance of 3DVWs’ application in evaluation in health care is the work of McCallum et al. Using qualitative evaluation, they explored students’ experiences of learning decision-making skills in the Second Life clinical simulation laboratory [29]. Table 7 shows the complete details of health care context and subcontexts of published papers related to evaluation [27-34].

Brief descriptions for each paper from the extracted papers in this category are shown in Table 8. For instance, the effects on individual performance of dyadic versus individual practice in a 3DVW home safety assessment were tested by Andrade et al [32]. Schwaab et al explored the use of Second Life virtual simulation technology to administer mock oral examinations to emergency medicine residents [34].

Table 7. Health care contexts and subcontexts in the evaluation category.

Context	Subcontext/Reference
Nursing	Decision-making skills [29]
Surgery	Different surgical scenarios [27]
Psychology	Mental health issues [30]
Safety	Elderly persons’ home safety assessment [31]
	Geriatric home safety simulation [32]
	Geriatric home safety [28]
Regulatory focus	Trustworthiness [33]
Emergency	Mock oral examination [34]

Table 8. Brief descriptions of papers in the evaluation category.

Subcontext	Brief description/quote
Decision-making skill	"… to explore nursing students' decision-making skills through the use of a 3D virtual environment such as Second Life" p. 699 [29].
Different surgical scenarios	Second Life was used for simulating different surgical scenarios by 63 surgeons for training and assessment. [27]
Mental health issues	"This paper describes the development of a 'rules discovery' type game [based on Second Life] for teaching construction sequencing and the ongoing efforts to generalize the rules-discovery framework for mental health remote assessment and wider educational use" p. 14 [30].
Elderly persons' home safety assessment	"… developed a 3-D home simulation in the virtual world Second Life containing 50 safety hazards that could affect the safety of an elderly person at home [for home safety assessment.]" p. 541 [31].
Geriatric home safety simulation	"This pilot study tests the effects on individual performance of dyadic versus individual practice in a 3D virtual world (VW) home safety assessment" p. 1 [32].
Geriatric home safety	"[Author's] aim was to evaluate the feasibility, usability, and acceptability of virtual worlds for geriatric home safety assessments and to correlate performance efficiency in hazard identification with spatial ability, self-efficacy, cognitive load, and presence" pp. 233-234 [28].
Trustworthiness	"This study examined the influence of regulatory focus and medical recommendation avatars' trustworthiness in avatar-based e-health within 3D virtual environments" p. 461 [33].
Mock oral examination	"In this study, [authors] explored the use of [Second Life] virtual simulation technology to administer mock oral examinations to emergency medicine residents" p. 559 [34].

Lifestyle

In the last decade, obesity, and overweight have become a global problem. According to our studies, in order to improve people's lifestyle, 3DVWs have been used in three different fitness programs. All three studies have used these worlds for obesity prevention and weight loss programs. For example, Johnston et al applied Second Life to implement a VW weight loss program for 54 overweight people [35]. In another study, Second Life was used by Siddiqi et al to implement an obesity prevention project [36]. See Table 9 for contexts and subcontexts of the studies using 3DVWs for lifestyle promotion [35-37].

Brief descriptions of these three papers from the extracted studies are shown in Table 10 to provide meaningful insights into each application. For instance, 3DVWs and their usefulness in obesity and diabetes therapies have been discussed by Morie and Chance [37].

Table 9. Health care contexts and subcontexts in the lifestyle category.

Context	Subcontext/Reference
Fitness	Obesity prevention [36]
	Obesity and diabetes/stress [37]
	Weight loss (behavioral change and self-efficacy) [35]

Table 10. Brief descriptions of papers in the lifestyle category.

Subcontext	Brief description
Obesity prevention	Implementation of an obesity prevention project using Second Life [36]
Obesity and diabetes/stress	How 3DVWs can be potentially powerful complements to obesity and diabetes therapies [37]
Weight loss (behavioral change and self-efficacy)	Efficacy of Second Life in a weight program relative to weight loss and behavioral change [35]

Academic Education

It is clear that 3DVWs are an emerging medium used in both traditional classrooms and distance education. In this study, we found the majority of studies applied 3DVWs for pedagogical purposes in health care. The education category consists of two main subcategories: academic education and professional education.

Academic education focuses on academic and university-related applications of 3DVWs in health care. This category is mostly related to training programs for students and staff in universities and educational communities. Papers in this category used 3DVWs in contexts such as nursing, emergency health, and public health. For instance, in 2012, Chow et al described the development and evaluation of 3DVW for learning rapid sequence intubation (RSI) by 206 nursing students [38]. Afterwards, they explored the intention of students to use the implemented system by means of the technology acceptance model (TAM) [39,40]. Second Life has been used by the University of Michigan School of Nursing to design and implement a virtual hospital to run virtual simulations for students [41]. Veronin and colleagues' work is another example of the use of 3DVWs in academic education. They developed an elective course at the Rangel College of Pharmacy for second- and third-year students [42]. Table 11 shows the contexts and subcontexts of the relevant papers [4,38,41-61].

Brief descriptions for each paper in the academic education category are shown in Table 12. For instance, Honey et al used Second Life for teaching postpartum hemorrhage to undergraduate nursing students and lecturers [48]. A development of virtual patient simulations for medical education using 3DVWs was performed by Danforth et al [57]. Another example of the application of 3DVWs in academic education is the work of Toth-Cohen and Gallagher [61]. They developed and evaluated a public exhibition on health and wellness at the Jefferson Occupational Therapy Education Center in Second Life.

Table 11. Health care contexts and subcontexts in the academic education category.

Context	Subcontext/reference
Nursing	Public health services [43]
	Graduate, undergraduate, and doctoral courses [44]
	Patient safety simulation [41]
	Nutrition [45]
	Public health issues (various activities) [46]
	Clinical simulation [47]
	Postpartum hemorrhage simulation [48]
	Acute-care medicine (emergency) [49]
	Disaster scenario [50]
	ICU first hour shift handover process [51]
	Rapid sequence intubation (RSI) [38]
Emergency	Accident rescue procedure [52]
	Trauma management [53]
	Speed and accuracy of nurse response [54]
	Pre-hospital CPR training [55]
	CPR [56]
	Managing emergencies and disasters [4]
Medical education	Virtual patient simulation [57]
Safety	Construction safety [58]
Health care higher education	Distance education [59]
Pharmacy	Drug safety [42]
Home assessment	Patient-centered decision-making [60]
Public health	Public health [61]

Table 12. Brief descriptions of papers in the academic education category.

Subcontext	Brief description/quote
Public health services	"…the authors describe how Second Life was integrated into a community nursing course" p. 74 [43].
Graduate, undergraduate, and doctoral courses	Second Life, as a 3DVW, found to be an environment that can provide valuable educational experiences in nursing [44].
Patient safety simulation	"The purpose of this article is to discuss how the University of Michigan School of Nursing designed and implemented a virtual hospital unit in Second Life to run virtual simulations" [41] p. 469.
Nutrition	Use of Second Life to teach interview skills [45].
Public health issues (various activities)	(1) Using Second Life in an online BSN program, (2) providing clinical experiences not often encountered, (3) using technology to help students feel connected to their classmates and instructors [46].
Clinical simulation	"… to explore Second Life as a clinical simulation platform, based on the attitudes and experiences of a sample of undergraduate nursing students" p. 883 [47].
Postpartum hemorrhage simulation	Use of Second Life for teaching postpartum hemorrhage to undergraduate nursing students and lecturers from New Zealand and the United States [48].
Acute-care medicine (emergency)	"[Authors] present three virtual world studies for team training and assessment in acute-care medicine: (1) training emergency department (ED) teams to manage individual trauma cases; (2) prehospital and in-hospital disaster preparedness training; (3) training ED and hospital staff to manage mass casualties after chemical, biological, radiological, nuclear, or explosive incidents" p. 161 [49].
Disaster scenario	"[Second Life]was implemented into an accelerated online nursing program" p. 152 [50].
ICU first hour shift handover process	"…aim was to design and develop a novel virtual world application for teaching and training Intensive Care nurses in the approach and method for shift handover, to provide an independent, but rigorous approach to teaching these important skills" p. 178 [51].
Rapid sequence intubation (RSI)	"…describes the development and evaluation of a virtual environment, the online 3D world Second Life (SL), for learning rapid sequence intubation (RSI)" p. 1136 [38].
Accident rescue procedure	"…describe experience developing virtual world-based training systems for two health care contexts. In one, procedural training was emphasised, while the other focused on teaching communication skills" p. 89 [52].
Trauma management	"…describes a project to develop and evaluate a computer-based simulator (the Virtual Emergency Department) for distance training in teamwork and leadership in trauma management" p. 321 [53].
Speed and accuracy of nurse response	"…explores immersive virtual reality as a potential educational strategy for nursing education and describes a project to develop and pioneer its use" p. 314 [54].
Pre-hospital CPR training	"…report on a study that investigates the relationship between repeated training of teams managing a medical emergency (CPR) in a Virtual World and performance outcome measures in a group of 12 medical students" p. 89 [55].
Cardiopulmonary resuscitation (CPR)	"In addition to finding a feasible way to implement CPR training, authors' aim was to investigate how a serious game setting in a virtual world using avatars would influence medical students' subjective experiences as well as their retention of knowledge" p. 1 [56].
Managing emergencies and disasters	"…explore the geo-data display potential of virtual worlds and their likely convergence with mirror worlds in the context of the future 3-D Internet or Metaverse, and reflect on the potential of such technologies and their future possibilities, eg, their use to develop emergency/public health virtual situation rooms to effectively manage emergencies and disasters in real time" p. 1 [4].
Virtual patient simulation	"the development of virtual patient simulations for medical education. In order to simulate real patients with greatest fidelity, the virtual patients [were] controlled by artificial intelligence" p. 3 [57].
Construction safety	"This paper proposes the adoption of online 3D world Second Life (SL) platform which allows students to perform role-playing, dialogic learning, and social interaction for efficient and effective construction safety and health education" p. 1 [58].
Distance education	Distance learning program in health care higher education using Second Life [59].
Drug safety	"…an elective course at the Rangel College of Pharmacy in pharmacy case studies for second- and third-year Doctor of Pharmacy students using Second Life" p. 105 [42].
Patient-centered decision-making	"The purpose of this research was to better understand the utility of a Web-based virtual environment as a teaching tool to represent clinical assessment and interventions in the home environment" p. 199 [60].
Public health	"…the development and evaluation of public exhibits on health and wellness at the Jefferson Occupational Therapy Education Center in Second Life" p. 3 [61].

Professional Education

The second category of education is professional education. In this category, researchers used 3DVWs for training in professional health care such as training programs for nurses, physicians, hospital staff, and so on, in health care contexts such as nursing, public health, radiotherapy, and clinical medicine. All papers in this group conducted educational programs for non-academic learners. For example, Andrade et al studied the feasibility of using 3DVW in training medical trainees to deliver bad news to patients [62]. Kamel Boulos designed a sexual health project in Second Life to provide education about sexually transmitted infections, unintended pregnancy, and improvements in sexual relationships [63,64]. Table 13 shows the professional education category of 3DVWs applied in various health care contexts and subcontexts [62-72].

Brief descriptions for each paper in the professional education category are shown in Table 14. For example, the usage of Second Life in health care education and its ability to improve patient safety were explored by Lee and Berge [69]. Watson et al presented a framework that demonstrates how 3DVWs can be applied to meet the needs of patients with diabetes [70].

Table 13. Health care contexts and subcontexts in the professional education category.

Context	Subcontext/Reference
Radiotherapy	Cancer care [66]
Nursing	Facilitate nursing journal clubs [67]
Patient-centered communication	Counseling patients about colorectal cancer [68]
	Delivering bad news to patients [62]
Health care higher education	Patient safety [69]
Clinical medical	Diabetes care [70]
	Diabetes type 2 [71]
Laboratory	Animal care [72]
Public health	Sexual health [63,64]
	Bio-terrorism defence event [65]

Table 14. Brief descriptions of papers in the professional education category.

Subcontext	Brief description/quote
Cancer care	“[authors] propose a novel Web based e-learning application design approach that uses Discrete Event System Specification (DEVS) formalism to form and model the 3D virtual hospital Web based application for radiotherapy and cancer care treatment class of applications” p. 1 [66].
Facilitate nursing journal clubs	“The purpose of this mixed-methods pilot study was to explore the feasibility of using Second Life to conduct research and to describe nurses’ experiences in using Second Life to facilitate nursing journal clubs” p. 146 [67].
Counseling patients about colorectal cancer	“…aims were to (1) explore the feasibility, acceptability, and effectiveness of a virtual-world platform for delivering MI training designed for physicians and (2) pilot test instructional designs using SL for MI training” p. 77 [68].
Delivering bad news to patients	8 medical trainees viewed an avatar-mediated training in Second Life as an instructional method for learning how to deliver bad news to patients [62].
Patient safety	“…the usage of Second Life in health care education and its ability to improve patient safety” p. 17 [69].
Diabetes care	“…a framework that demonstrates how applications within SL can be constructed to meet the needs of patients with diabetes, allowing them to attend group visits, learn more about lifestyle changes, and foster a sense of support and emotional well-being” p. 697 [70].
Diabetes Type 2	“a pilot postgraduate medical education program in the virtual world, Second Life” p. 1 [71].
Animal care	“The article explores the potential utility of multi-user virtual environments for advancing laboratory animal care and use through better education and training” p. 163 [72].
Sexual health	“…an evaluation of a sexual health project in Second Life, designed to provide education about sexually transmitted infections, prevention of unintended pregnancy and promotion of equalitarian sexual relationships” p. 279 [63].
	“…the University of Plymouth Sexual Health SIM in Second Life… and provides some reflections on its design, as well as some details about the planned evaluation of the project” p. 1 [64].
Bio-terrorism defence event	Creation and evaluation of a pilot bioterrorism defence training environment using virtual reality technology [65].

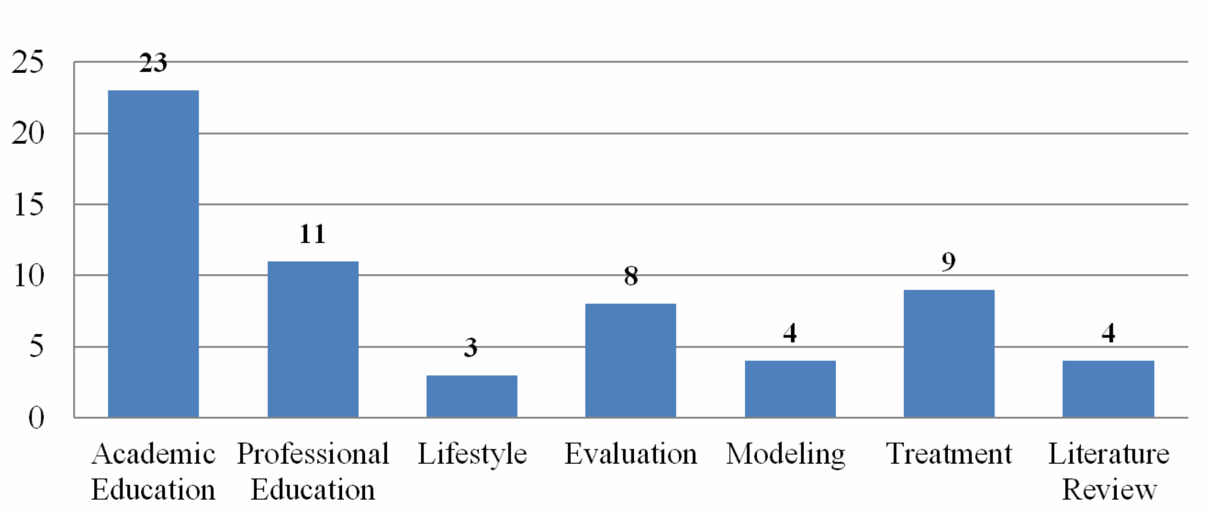
Discussion

Principal Findings

In this study, we attempted to provide an overview of 3DVWs and their application areas in health care contexts. To this end, we found 62 different papers from five popular scientific databases. To gain a general understanding of 3DVW research, we classified these 62 research studies into six meaningful categories. Figure 6 shows the number of published studies in each category. The education category including professional and academic education contains the largest number of papers (n=34), of which 23 are related to the academic education category and 11 to the professional education category. In contrast, three papers are related to the lifestyle category, which has the least number of articles. In spite of the fact that treatment plays a crucial role in health care, only nine papers are allocated to this category. Eight papers have content related to assessment and evaluation, and in four of the papers the authors used 3DVWs as a platform for modeling various health-related environments. Of the total 62 papers, four papers included surveys and literature reviews on this subject.

3DVWs are increasing in popularity as a new medium for educational purposes, and pedagogical institutions are adopting this technology to support their teaching and learning. A wide range of well-known educational institutions around the world are using 3DVWs for various purposes, including distance education, presentations, meetings, and literature and language attainment. Since 3DVWs do not have a storyline or plot of avatars, actors, and events, the lack of a guiding narrative in these technologies provides flexibility for educators to design more complicated learning spaces for their pedagogical requirements. Not only is the technology propeling the use of 3DVWs in education and learning, but the new generation of students and learners is also demanding the use of this advanced technology in education. These are the main reasons that education, especially academic education, attracted more attention in the health care sector, compared to the other application areas.

Figure 6. Number of published papers based on the categories.



Implications for Researchers

In spite of the large number of studies on 3DVWs in health care, no systematic review has been conducted on this topic. There are several research papers that survey the range of health-related activities in virtual worlds [7-10]. For example, Kamel Boulos et al, in their hybrid review/case study, introduced 3DVWs and the educational potential of Second Life to medical/health librarians and educators. They described medical and health education examples from Second Life and compiled a companion resources page [73] with additional online material to support their study [7]. Hansen et al, in her review of existing literature, described the potential of health care learning environments in 3DVWs and provided relevant examples of 3DVWs [8]. Beard et al performed a literature survey and categorized the range of health-related activities in Second Life. They identified 68 health-related sites and examined the design attributes of the sites and assessed the utilities of Second Life for propagation of health information. They developed five categories to explain the range of online health-related activities [10]. But those literature reviews do not provide a comprehensive analysis of the topic, and their surveys were published between 2007 and 2010. There was no systematic review with a sufficient number of reviewed articles. In order to update the literature and provide a more comprehensive picture of 3DVW applications in health care, we decided to perform a systematic review examining 62 papers, a larger number than that covered by previous surveys and literature reviews. We also covered publications from January 1990 to July 2013.

The proposed taxonomy provides a big picture of the application areas of 3DVWs in health care. The most reviewed papers in this systematic review are in the academic and professional education categories—almost 54% of all papers (34/62). Despite the fact that researchers have paid more attention to education, there are gaps in this category. For example, studying the impact of 3DWVs in the education of surgeons could be an excellent starting point. Creating specific rooms and environments in virtual worlds so students can remotely access their course materials such as files, e-books, lecture captures, and presentation slides is a second suggestion. Training students and the public in first aid with 3DVWs is another.

Living a healthy lifestyle has become a common ambition, and people are now attempting to pay more attention to their health. There are, however, few studies in the literature regarding the impact of 3DVWs on people's health and lifestyle. Therefore, investigating the advantages and disadvantages of applying this new technology in improving healthy behaviors and extending the health culture could be a useful endeavor.

Although there are studies related to treatment in health care and medicine in the current literature, there is a lack of research on some application areas of 3DVWs in this category. For example, in the treatment and psychology context, no study addresses problems such as social isolation, care of the elderly, phobias, and so on.

Existing studies pay little attention to the application of 3DVWs in modeling. According to our systematic review, few have been conducted in this field. Therefore, the powerful ability of these tools for modeling and simulating health scenarios could be investigated in greater depth. For example, investigating the impact of simulated sophisticated hospital equipment on the skill-building of hospital staff would be worth examining. Future research could also focus on replication of earlier studies with larger sample sizes to examine the impact of 3DVWs on various health care contexts.

Implications for Practice

Our systematic review and the proposed taxonomy have various implications for a wide range of individuals and institutions that could use this study to recognize different areas of application of 3DVWs of benefit to their professional, business, and organizational decisions, for example, health decision-makers, universities, educational communities, hospitals, physicians and practitioners, nurses, patients, nursing homes, pharmaceutical institutes, health marketers, disaster managers, psychologists, public health providers, clinical medicine, addiction treatment institutes, etc.

Our systematic review shows that pedagogical uses are the most important applications of 3DVWs in health care. Universities, hospitals, health care organizations, training centers, educational institutions, and educators can use 3DVWs as a creative, powerful, and efficient tool for courses, lectures, and workshops. They can develop new effective ways of teaching and learning in virtual environments. Thanks to the Internet-based characteristics of these worlds, there is the possibility of distance learning; therefore, students, nurses, physicians, medical staff, and other trainees can connect to the specified worlds and attend a virtual classroom using their own computer. Teachers and trainers also have wider access to geographically remote trainees and distance learners.

Universities, colleges, hospitals, health care agencies, private groups, businesses, and corporations can use 3DVWs for discussion, seminars, presentations, meetings, and other similar activities by creating large virtual auditoriums and meeting rooms, and gathering their staff in specific environments within virtual worlds where digital materials can be created, stored, and used. These communities could also create virtual health and medical libraries to enable remote access to related e-books, documents, and other applications; 3DVWs have great potential for librarians and educators in remote locations.

We found that hospitals, universities, and medical training organizations can model laboratories, wards, and various parts of a hospital with instruments, equipment, and settings that simulate real-world activities at a very low cost for educational and pedagogical purposes. Virtual simulations in 3DVWs can be an acceptable strategy for delivering scenarios that focus on health-related skills. The critical life simulation enables educators to collaborate and solve problem-based scenarios in a team, allowing them to construct personal and technical skills actively through interaction in a virtual environment without the danger of making mistakes and risking harm to patients or themselves.

Our findings could benefit psychologists and psychological institutions with use of 3DVWs in their treatment and therapeutic measurements. These worlds could be an interesting tool for prevention and treatment of different psychological disorders. People with psychological and mental problems could benefit from using these worlds under the supervision of expert medical staff. Mental health specialists could devise specific remedial programs for patients suffering from psychological stress, anxiety, PTSD, schizophrenia, and other problems.

Addiction treatment institutes and therapeutic centres could use 3DVWs in their therapeutic programs to reduce the risk of abuse and addiction. Patients suffering from addiction disorders could receive behavioral treatments in virtual worlds that include the planning of specific ways to avoid addictive stimulus. By using therapeutic interventions in 3DVWs, addiction medicine specialists may be able to help patients learn healthier ways to find satisfaction.

Disaster managers could use 3DVWs for disaster preparedness and response training programs. By simulating similar disaster scenarios, they would be able to increase the experience, collaboration, leadership, disaster response, and decision-making skills of their personnel in various disaster scenarios.

Nursing homes and institutes with a responsibility for aged people might be able to use these worlds as an interesting tool for treatment as well as prevention of conditions such as social isolation. They could also be used to entertain people with physical disabilities and help them to cope with loneliness and social isolation.

It is clear that 3DVWs offer a wide range of features that could enhance marketing companies' and health-related organizations' marketing initiatives. Traditional advertising is currently replicated in VWs with digital billboard advertisements and product placement. Current and future real-life initiatives could also be replicated in these platforms as different simulations. Communities could use these platforms as a way of promoting hospitals, health care services, health system reform, and even fundraising.

Health decision-makers and policy-makers, health IT managers, health marketers, and medical business owners could use these results to find different areas of application of 3DVWs of benefit to their organizations. Furthermore, public health providers could focus on 3DVWs to extend public awareness and promote healthy lifestyles. For example, they could include videos, presentations, buildings, and even exercise and sport facilities in their private virtual islands to motivate users to adopt healthier behaviors in their daily life.

Patients with different kinds of social disorders and people who have problems communicating with others for different reasons could use 3DVWs to improve their social learning and interactive behaviors. A 3D virtual experience gives patients a feeling of control over their health, improving their knowledge and confidence, since they can navigate the health care system from their own home.

Physicians and practitioners can be aware of the latest virtual facilities in different 3DVWs and apply them in their therapeutic and medical treatments. They can create specific environments to meet their clinical needs as well as the needs of their patients. For example, they can communicate with their patients through these platforms to perform check-ups and issue prescriptions and necessary guidance.

3DVWs play a crucial role in the assessment and evaluation of diverse skills in health care. For example, virtual patients can be used by trainees such as nurses, surgeons, students, and other medical staff, and their performance can be assessed and benchmarked in different ways.

Limitations

The quality of these results is highly dependent on the quality of the papers that have been identified. Therefore, it is not possible to evaluate the quality of the results in this study. We cannot guarantee that we have taken all relevant applications of 3DVWs in health care into account because there are limited studies related to this topic. Any systematic review is limited to its keywords; this study is no exception. We cannot present any global conclusions about the application areas of 3DVWs in health care because only a few of the studies were conducted in countries other than the United States and the United Kingdom.

Conclusions

This study provides a comprehensive picture of 3DVW’s application in health care and updates the literature in this field. It also highlights various health contexts and subcontexts that have applied 3DVWs. The results of this systematic literature review could be beneficial for researchers interested in this topic to better understand the field and previous studies, better classify research, help them shape the future direction of research, and identify gaps in the literature. By updating work in the application areas of 3DVWs, this study also attempts to help a wide variety of individuals and organizations, such as practitioners, nurses, managers, hospitals, health care agencies, private groups, business health companies and corporations, and universities to recognize various areas of application of 3DVWs and determine directions for practice in their own areas.

To conclude, it is evident that 3D virtual worlds present several innovative ways to carry out health-related activities. In this study, we developed six main categories to explain the application areas of 3DVWs in various health care contexts: academic education, professional education, treatment, modeling, lifestyle, and evaluation. Our proposed taxonomy could be used to provide an overview of the application of 3DVWs in health care and medical research and practice that individuals, professional health communities, and academic institutions could use in their various activities.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms used in each database.

[PDF File (Adobe PDF File), 248KB - jmir_v16i2e47_app1.pdf]

References

1. 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]
2. 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]
3. Bell MW. Toward a definition of "virtual worlds". Journal of Virtual Worlds Research 2008;1(1):1-5 [FREE Full text] [doi: 10.4101/jvwr.v1i1.283]
4. Boulos MN, Burden D. Web GIS in practice V: 3-D interactive and real-time mapping in Second Life. Int J Health Geogr 2007;6:51 [FREE Full text] [doi: 10.1186/1476-072X-6-51] [Medline: 18042275]
5. Second Life. URL: http://secondlife.com/ [accessed 2014-01-03] [WebCite Cache ID 6MLKBplM1]
6. SL10B Infographic: Ten Years of Second Life. 2014-01-03 URL: http://community.secondlife.com/t5/Featured-News/SL10B-Infographic-Ten-Years-of-Second-Life/ba-p/2053857 [accessed 2014-01-03] [WebCite Cache ID 6MLKPKzib]
7. Boulos MN, Hetherington L, Wheeler S. Second Life: an overview of the potential of 3-D virtual worlds in medical and health education. Health Info Libr J 2007 Dec;24(4):233-245. [doi: 10.1111/j.1471-1842.2007.00733.x] [Medline: 18005298]

8. 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]
9. Peck B, Miller C. I think I can, I think I can, I think I can…I know I can Multi-user Virtual Environments (MUVEs) as a means of developing competence and confidence in undergraduate nursing students An Australian perspective. Procedia - Social and Behavioral Sciences 2010 Jan;2(2):4571-4575. [doi: 10.1016/j.sbspro.2010.03.732]
10. 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]
11. Kitchenham B. Procedures for performing systematic reviews. Keele University and National ICT Australia Ltd 2004;2004:1-28.
12. Ghapanchi AH, Aurum A. Antecedents to IT personnel's intentions to leave: A systematic literature review. Journal of Systems and Software 2011 Feb;84(2):238-249. [doi: 10.1016/j.jss.2010.09.022]
13. Bakshi A, Talaei-Khoei A, Ray P. Adaptive policy framework: A systematic review. Journal of Network and Computer Applications 2013 Jul;36(4):1261-1271. [doi: 10.1016/j.jnca.2012.12.007]
14. 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]
15. Ku J, Han K, Lee HR, Jang HJ, Kim KU, Park SH, et al. VR-based conversation training program for patients with schizophrenia: a preliminary clinical trial. Cyberpsychol Behav 2007 Aug;10(4):567-574. [doi: 10.1089/cpb.2007.9989] [Medline: 17711366]
16. Hoch DB, Watson AJ, Linton DA, Bello HE, Senelly M, Milik MT, et al. The feasibility and impact of delivering a mind-body intervention in a virtual world. PLoS One 2012;7(3):e33843 [FREE Full text] [doi: 10.1371/journal.pone.0033843] [Medline: 22470483]
17. Pioggia G, Carbonaro N, Anania G, Tognetti A, Tartarisco G, Ferro M, et al. Interreality: The use of advanced technologies in the assessment and treatment of psychological stress. In: IEEE. 2010 Presented at: 10th International Conference on Intelligent Systems Design and Applications (ISDA); Nov. 29-Dec. 1, 2010; Cairo p. 1047-1051. [doi: 10.1109/ISDA.2010.5687047]
18. Kusumi T, Ogura K, Miura A. The Development of a Positive Community using Virtual Space for Cancer Patients. In: ISUC '08. 2008 Presented at: Second International Symposium on Universal Communication; Dec 15-16, 2008; Osaka p. 490-493. [doi: 10.1109/ISUC.2008.48]
19. Dayan E. ARGAMAN: Rapid Deployment Virtual Reality System for PTSD Rehabilitation. In: ITRE '06. 2006 Presented at: International Conference on Information Technology: Research and Education; Oct. 16-19, 2006; Tel-Aviv p. 34-38. [doi: 10.1109/ITRE.2006.381528]
20. Rizzo AA, Lange B, Buckwalter JG, Forbell E, Kim J, Sagae K, et al. An intelligent virtual human system for providing healthcare information and support. Stud Health Technol Inform 2011;163:503-509. [Medline: 21335847]
21. Riva G, Raspelli S, Algeri D, Pallavicini F, Gorini A, Wiederhold BK, et al. Interreality in practice: bridging virtual and real worlds in the treatment of posttraumatic stress disorders. Cyberpsychol Behav Soc Netw 2010 Feb;13(1):55-65. [Medline: 20528294]
22. Gustafson D, Shaw B, Isham A, Dillon D, Spartz J. Robert Wood Johnson Foundation. 2008. Exploring the potential of the web-based virtual world of Second Life to improve substance abuse treatment outcomes URL: https://www.niatx.net/pdf/enews/september2008/secondlife_report.pdf [WebCite Cache ID 6NH3asg6c]
23. Kamel Boulos MN, Ramloll R, Jones R, Toth-Cohen S. Web 3D for public, environmental and occupational health: early examples from Second Life. Int J Environ Res Public Health 2008 Dec;5(4):290-317 [FREE Full text] [Medline: 19190358]
24. Thompson CW, Hagstrom F. Modeling Healthcare Logistics in a Virtual World. IEEE Internet Comput 2008 Sep;12(5):100-104. [doi: 10.1109/MIC.2008.106]
25. Jin SA, Lee KM. The influence of regulatory fit and interactivity on brand satisfaction and trust in E-health marketing inside 3D virtual worlds (Second Life). Cyberpsychol Behav Soc Netw 2010 Dec;13(6):673-680. [doi: 10.1089/cyber.2009.0292] [Medline: 21142992]
26. Boulos MN, Scotch M, Cheung KH, Burden D. Web GIS in practice VI: a demo playlist of geo-mashups for public health neogeographers. Int J Health Geogr 2008;7:38 [FREE Full text] [doi: 10.1186/1476-072X-7-38] [Medline: 18638385]
27. Patel V, Aggarwal R, Cohen D, Taylor D, Darzi A. Implementation of an interactive virtual-world simulation for structured surgeon assessment of clinical scenarios. J Am Coll Surg 2013 Aug;217(2):270-279. [doi: 10.1016/j.jamcollsurg.2013.03.023] [Medline: 23870219]
28. Andrade AD, Cifuentes P, Mintzer MJ, Roos BA, Anam R, Ruiz JG. Simulating geriatric home safety assessments in a three-dimensional virtual world. Gerontol Geriatr Educ 2012;33(3):233-252. [doi: 10.1080/02701960.2011.611553] [Medline: 22816973]
29. McCallum J, Ness V, Price T. Exploring nursing students' decision-making skills whilst in a Second Life clinical simulation laboratory. Nurse Educ Today 2011 Oct;31(7):699-704. [doi: 10.1016/j.nedt.2010.03.010] [Medline: 20807671]
30. Strelzoff A, Sulbaran T. Transformation of a Collaborative Virtual Reality Environment for construction scheduling to help individual with mental health issues. In: FIE 2008. 2008 Presented at: 38th Annual Frontiers in Education Conference; Oct.22-25, 2008; Saratoga Springs, NY p. S4J-13-S4J-18. [doi: 10.1109/FIE.2008.4720671]

31. Andrade AD, Cifuentes P, Oliveira MC, Anam R, Roos BA, Ruiz JG. Avatar-mediated home safety assessments: piloting a virtual objective structured clinical examination station. J Grad Med Educ 2011 Dec;3(4):541-545 [FREE Full text] [doi: 10.4300/JGME-D-11-00236.1] [Medline: 23205205]
32. Andrade AD, Anam R, Sun H, Mintzer MJ, Ruiz JG. Effects on performance of individual versus dyadic practice during an avatar-based three-dimensional virtual home safety simulation. Stud Health Technol Inform 2013;184:13-19. [Medline: 23400122]
33. Jin SA. The roles of regulatory focus and medical recommendation avatars' trustworthiness in virtual environment-based e-health. Cyberpsychol Behav Soc Netw 2010 Aug;13(4):461-466. [doi: 10.1089/cyber.2009.0291] [Medline: 20712505]
34. Schwaab J, Kman N, Nagel R, Bahner D, Martin DR, Khandelwal S, et al. Using Second Life virtual simulation environment for mock oral emergency medicine examination. Acad Emerg Med 2011 May;18(5):559-562. [doi: 10.1111/j.1553-2712.2011.01064.x] [Medline: 21521404]
35. Johnston JD, Massey AP, Devaneaux CA. Innovation in weight loss programs: a 3-dimensional virtual-world approach. J Med Internet Res 2012;14(5):e120 [FREE Full text] [doi: 10.2196/jmir.2254] [Medline: 22995535]
36. Siddiqi S, Mama SK, Lee RE. Developing an obesity prevention intervention in virtual worlds: The international health challenge in Second Life. Journal of Virtual Worlds Research 2011;3(3):3-26 [FREE Full text] [doi: 10.4101/jvwr.v3i3.809]
37. Morie JF, Chance E. Extending the reach of health care for obesity and diabetes using virtual worlds. J Diabetes Sci Technol 2011 Mar;5(2):272-276 [FREE Full text] [Medline: 21527093]
38. Chow M, Herold DK, Choo TM, Chan K. Extending the technology acceptance model to explore the intention to use Second Life for enhancing healthcare education. Computers & Education 2012 Dec;59(4):1136-1144. [doi: 10.1016/j.compedu.2012.05.011]
39. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
40. Davis FD. A technology acceptance model for empirically testing new end-user information systems: Theory and results. In: Massachusetts Institute of Technology. Boston: Unpublished doctoral dissertation; 1986.
41. Aebersold M, Tschannen D, Stephens M, Anderson P, Lei X. Second Life: A new strategy in educating nursing students. Clinical Simulation in Nursing 2012 Nov;8(9):e469-e475. [doi: 10.1016/j.ecns.2011.05.002]
42. Veronin MA, Daniels L, Demps E. Pharmacy cases in Second Life: an elective course. Adv Med Educ Pract 2012;3:105-112 [FREE Full text] [doi: 10.2147/AMEP.S35358] [Medline: 23762008]
43. Schmidt B, Stewart S. Implementing the virtual world of Second Life into community nursing theory and clinical courses. Nurse Educ 2010;35(2):74-78. [doi: 10.1097/NNE.0b013e3181ced999] [Medline: 20173592]
44. Warren JJ, Connors HR, Trangenstein PA. A paradigm shift in simulation: Experiential learning in Second Life. In: Saba V, McCormick K, editors. Essentials of Nursing Informatics. New York: McGraw Hill; 2011:619-631.
45. Sweigart L, Hodson-Carlton K, Campbell BR, Lutz DR. Second life environment: a venue for interview skill development. Comput Inform Nurs 2010;28(5):258-263. [doi: 10.1097/01.NCN.0000388349.67292.3c] [Medline: 20736722]
46. Stewart S, Hansen TS, Pope D, Schmidt B, Thyes JG, Jambunathan J, et al. Developing a Second Life campus for online accelerated BSN students. Comput Inform Nurs 2010;28(5):253-258. [doi: 10.1097/01.NCN.0000388348.29174.9f] [Medline: 20736721]
47. Rogers L. Simulating clinical experience: Exploring Second Life as a learning tool for nurse education. In: 26th Annual ASCILITE International Conference. 2009 Presented at: Auckland, Same places, different spaces; Dec 6-9, 2009; Auckland, New Zealand p. 883-887 URL: http://ascilite.org.au/conferences/auckland09/procs/rogers.pdf
48. Honey M, Diener S, Connor K, Veltman M, Bodily D. Teaching in virtual space: Second Life simulation for haemorrhage management. In: 26th Annual ASCILITE International Conference. 2009 Presented at: Auckland, Same places, different spaces; Dec 6-9, 2009; Auckland, New Zealand p. 1222-1224 URL: http://ascilite.org.au/conferences/auckland09/procs/honey-interactive-session.pdf
49. LeRoy Heinrichs W, Youngblood P, Harter PM, Dev P. Simulation for team training and assessment: case studies of online training with virtual worlds. World J Surg 2008 Feb;32(2):161-170. [doi: 10.1007/s00268-007-9354-2] [Medline: 18188640]
50. Schmidt B, Stewart S. Implementing the virtual reality learning environment: Second Life. Nurse Educ 2009;34(4):152-155. [doi: 10.1097/NNE.0b013e3181aabbe8] [Medline: 19574850]
51. Brown R, Rasmussen R, Baldwin I, Wyeth P. Design and implementation of a virtual world training simulation of ICU first hour handover processes. Aust Crit Care 2012 Aug;25(3):178-187. [doi: 10.1016/j.aucc.2012.02.005] [Medline: 22436543]
52. Chodos D, Stroulia E, Boechler P, King S, Kuras P, Carbonaro M, et al. Healthcare education with virtual-world simulations. In: SEHC '10. 2010 Presented at: ICSE Workshop on Software Engineering in Health Care; May 3-4, 2010; Cape Town, South Africa p. 89-99. [doi: 10.1145/1809085.1809097]
53. Dev P, Youngblood P, Heinrichs WL, Kusumoto L. Virtual worlds and team training. Anesthesiol Clin 2007 Jun;25(2):321-336. [doi: 10.1016/j.anclin.2007.03.001] [Medline: 17574193]
54. Kilmon CA, Brown L, Ghosh S, Mikitiuk A. Immersive virtual reality simulations in nursing education. Nurs Educ Perspect 2010;31(5):314-317. [Medline: 21086871]

55. Creutzfeldt J, Hedman L, Medin C, Wallin CJ, Felländer-Tsai L. Effects of repeated CPR training in virtual worlds on medical students' performance. Stud Health Technol Inform 2008;132:89-94. [Medline: 18391263]
56. Creutzfeldt J, Hedman L, Medin C, Heinrichs WL, Felländer-Tsai L. Exploring virtual worlds for scenario-based repeated team training of cardiopulmonary resuscitation in medical students. J Med Internet Res 2010;12(3):e38 [FREE Full text] [doi: 10.2196/jmir.1426] [Medline: 20813717]
57. Danforth DR, Procter M, Chen R, Johnson M, Heller R. Development of virtual patient simulations for medical education. Journal of Virtual Worlds Research 2009;2(2):3-11 [FREE Full text] [doi: 10.4101/jvwr.v2i2.707]
58. Park CS, Le QT. Construction safety education model based on second life. In: IEEE. 2012 Presented at: IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE); Aug. 20-23, 2012; Hong Kong p. H2C-1-H2C-5. [doi: 10.1109/TALE.2012.6360336]
59. Johnson CM, Vorderstrasse AA, Shaw R. Virtual worlds in health care higher education. Journal of Virtual Worlds Research 2009;2(2):3-12 [FREE Full text] [doi: 10.4101/jvwr.v2i2.699]
60. Sabus C, Sabata D, Antonacci D. Use of a virtual environment to facilitate instruction of an interprofessional home assessment. J Allied Health 2011;40(4):199-205. [Medline: 22138875]
61. Toth-Cohen S, Gallagher T. Development and evaluation of health and wellness exhibits at the Jefferson Occupational Therapy Education Center in Second Life. Journal of Virtual Worlds Research 2009;2(2):3-13 [FREE Full text] [doi: 10.4101/jvwr.v2i2.708]
62. Andrade AD, Bagri A, Zaw K, Roos BA, Ruiz JG. Avatar-mediated training in the delivery of bad news in a virtual world. J Palliat Med 2010 Dec;13(12):1415-1419. [doi: 10.1089/jpm.2010.0108] [Medline: 21091407]
63. Kamel Boulos MN, Toth-Cohen S. The University of Plymouth Sexual Health SIM experience in Second Life: evaluation and reflections after 1 year. Health Info Libr J 2009 Dec;26(4):279-288. [doi: 10.1111/j.1471-1842.2008.00831.x] [Medline: 19930475]
64. Kamel Boulos MN, Toth-Cohen S. Design for learning in 3-D virtual worlds: The University of Plymouth Sexual Health SIM experience in Second Life. University of Plymouth 2007;2007:1-6 [FREE Full text]
65. Yellowlees P, Cook JN, Marks SL, Wolfe D, Mangin E. Can virtual reality be used to conduct mass prophylaxis clinic training? A pilot program. Biosecur Bioterror 2008 Mar;6(1):36-44. [doi: 10.1089/bsp.2007.0038] [Medline: 18386971]
66. Al Hamidi A, Boukerche A, Ahmad L, Zhang M. Supporting web-based e-learning through collaborative virtual environments for radiotherapy treatment: A formal design. In: VECIMS 2008. 2008 Presented at: Virtual Environments, Human-Computer Interfaces and Measurement Systems; July 14-16, 2008; Istanbul p. 51-56. [doi: 10.1109/VECIMS.2008.4592752]
67. Billingsley L, Rice K, Bennett M, Thibeau S. Using a multiuser virtual environment to facilitate nursing journal clubs: a mixed-methods study. Clin Nurse Spec 2013;27(3):146-154. [doi: 10.1097/NUR.0b013e31828c8408] [Medline: 23575171]
68. Mitchell S, Heyden R, Heyden N, Schroy P, Andrew S, Sadikova E, et al. A pilot study of motivational interviewing training in a virtual world. J Med Internet Res 2011;13(3):e77 [FREE Full text] [doi: 10.2196/jmir.1825] [Medline: 21946183]
69. Lee A, Berge ZL. Second Life in Healthcare Education: Virtual Environment's Potential to Improve Patient Safety. Knowledge Management & E-Learning: An International Journal (KM&EL) 2011;3(1):17-23 [FREE Full text]
70. Watson AJ, Grant RW, Bello H, Hoch DB. Brave new worlds: how virtual environments can augment traditional care in the management of diabetes. J Diabetes Sci Technol 2008 Jul;2(4):697-702 [FREE Full text] [Medline: 19885247]
71. Wiecha J, Heyden R, Sternthal E, Merialdi M. Learning in a virtual world: experience with using second life for medical education. J Med Internet Res 2010;12(1):e1 [FREE Full text] [doi: 10.2196/jmir.1337] [Medline: 20097652]
72. Ketelhut DJ, Niemi SM. Emerging technologies in education and training: applications for the laboratory animal science community. ILAR J 2007;48(2):163-169 [FREE Full text] [Medline: 17420537]
73. HealthCyberMap. 2014-01-03 URL: http://healthcybermap.org/sl.htm [accessed 2014-01-03] [WebCite Cache ID 6MLKS5cTa]

Abbreviations

3D: three dimensional
3DVW: three-dimensional virtual world
CPR: cardiopulmonary resuscitation
PTSD: posttraumatic stress disorder
RFID: radio-frequency identification
RSI: rapid sequence intubation
TAM: technology acceptance model
VR: virtual reality
VW: virtual world

Edited by G Eysenbach; submitted 11.11.13; peer-reviewed by M Fasanghari, F Bidar; comments to author 26.12.13; revised version received 06.01.14; accepted 10.01.14; published 18.02.14.

Please cite as:
Ghanbarzadeh R, Ghapanchi AH, Blumenstein M, Talaei-Khoei A
A Decade of Research on the Use of Three-Dimensional Virtual Worlds in Health Care: A Systematic Literature Review
J Med Internet Res 2014;16(2):e47
URL: http://www.jmir.org/2014/2/e47/
doi:10.2196/jmir.3097
PMID:24550130

Original Paper

Evaluation of a Novel Conjunctive Exploratory Navigation Interface for Consumer Health Information: A Crowdsourced Comparative Study

Licong Cui[1], MS; Rebecca Carter[2], BA; Guo-Qiang Zhang[1], PhD

[1]Department of Electrical Engineering and Computer Science, Division of Medical Informatics, Case Western Reserve University, Cleveland, OH, United States

[2]School of Medicine, Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, United States

Corresponding Author:
Guo-Qiang Zhang, PhD
Department of Electrical Engineering and Computer Science
Division of Medical Informatics
Case Western Reserve University
2103 Cornell Road
Cleveland, OH, 44106
United States
Phone: 1 216 368 3286
Fax: 1 216 368 0207
Email: gq@case.edu

Abstract

Background: Numerous consumer health information websites have been developed to provide consumers access to health information. However, lookup search is insufficient for consumers to take full advantage of these rich public information resources. Exploratory search is considered a promising complementary mechanism, but its efficacy has never before been rigorously evaluated for consumer health information retrieval interfaces.

Objective: This study aims to (1) introduce a novel Conjunctive Exploratory Navigation Interface (CENI) for supporting effective consumer health information retrieval and navigation, and (2) evaluate the effectiveness of CENI through a search-interface comparative evaluation using crowdsourcing with Amazon Mechanical Turk (AMT).

Methods: We collected over 60,000 consumer health questions from NetWellness, one of the first consumer health websites to provide high-quality health information. We designed and developed a novel conjunctive exploratory navigation interface to explore NetWellness health questions with health topics as dynamic and searchable menus. To investigate the effectiveness of CENI, we developed a second interface with keyword-based search only. A crowdsourcing comparative study was carefully designed to compare three search modes of interest: (A) the topic-navigation-based CENI, (B) the keyword-based lookup interface, and (C) either the most commonly available lookup search interface with Google, or the resident advanced search offered by NetWellness. To compare the effectiveness of the three search modes, 9 search tasks were designed with relevant health questions from NetWellness. Each task included a rating of difficulty level and questions for validating the quality of answers. Ninety anonymous and unique AMT workers were recruited as participants.

Results: Repeated-measures ANOVA analysis of the data showed the search modes A, B, and C had statistically significant differences among their levels of difficulty ($P<.001$). Wilcoxon signed-rank test (one-tailed) between A and B showed that A was significantly easier than B ($P<.001$). Paired t tests (one-tailed) between A and C showed A was significantly easier than C ($P<.001$). Participant responses on the preferred search modes showed that 47.8% (43/90) participants preferred A, 25.6% (23/90) preferred B, 24.4% (22/90) preferred C. Participant comments on the preferred search modes indicated that CENI was easy to use, provided better organization of health questions by topics, allowed users to narrow down to the most relevant contents quickly, and supported the exploratory navigation by non-experts or those unsure how to initiate their search.

Conclusions: We presented a novel conjunctive exploratory navigation interface for consumer health information retrieval and navigation. Crowdsourcing permitted a carefully designed comparative search-interface evaluation to be completed in a timely and cost-effective manner with a relatively large number of participants recruited anonymously. Accounting for possible biases, our study has shown for the first time with crowdsourcing that the combination of exploratory navigation and lookup search is more effective than lookup search alone.

(J Med Internet Res 2014;16(2):e45) doi:10.2196/jmir.3111

KEYWORDS

crowdsourcing; consumer health information; human computer interaction; information retrieval; search interfaces; comparative user evaluation

Introduction

The Internet has become one of the most important sources for consumers to seek health-related information. As a recent survey indicated [1], over 80% of Internet users in the United States have looked online for health information such as a specific disease or treatment, and 60% said the information found online affected decisions or actions on their own health or the way they cared for someone else.

Although a substantial amount of consumer health information is available online [2], it is not necessarily easy for general consumers to access such information. For example, a study by Berland et al [3] found that accessing health information by means of search engines (eg, Google or Yahoo) and simple search terms was not efficient. Less than a quarter of links on the search engine's first pages of search results led to relevant content.

There are two basic information retrieval modes for accessing online health information: *lookup* and *exploratory* searches [4,5]. In lookup mode, a user comes with specific terms about a disease, medication, or other related description, enters search terms into search engines, and tries to retrieve a corresponding set of responses. For example, Berland et al [3] used lookup for the retrieval of an initial set of Web links by entering search terms such as "breast cancer", "childhood asthma", "depression", and "obesity" into search engines. In exploratory mode, a user may not have a specific target, or cannot easily and effectively formulate descriptive lookup terms, and may rely on navigational menus or facets to browse and explore the content. In most cases [6-13], lookup is accompanied by exploration to help the user find a needle in a haystack. The volume of search results can be overwhelmingly large and needs to be further structured to allow relevant information to be located. For example, Mu et al [7] presented a facet-view interface complementing lookup search for effectively retrieving and navigating medical literatures in a subset of MEDLINE [14].

Navigational exploration relies on information organization to provide structures (eg, topics as menus) with which to organize a collection of contents to facilitate browsing and exploration. Consumer health questions online are often organized by categories or topics in consumer health-related Web services such as WebMD Answers [15], health category in Yahoo Answers [16], and Ask an Expert in NetWellness [17,18].

However, a common limitation of these organizational structures is that each question is assigned a single topic among a collection of available topics, even though multiple topics are related to the question. This presents a major impediment to accessing consumer health information through use of navigational exploration, such as in searching the health question repository in NetWellness.

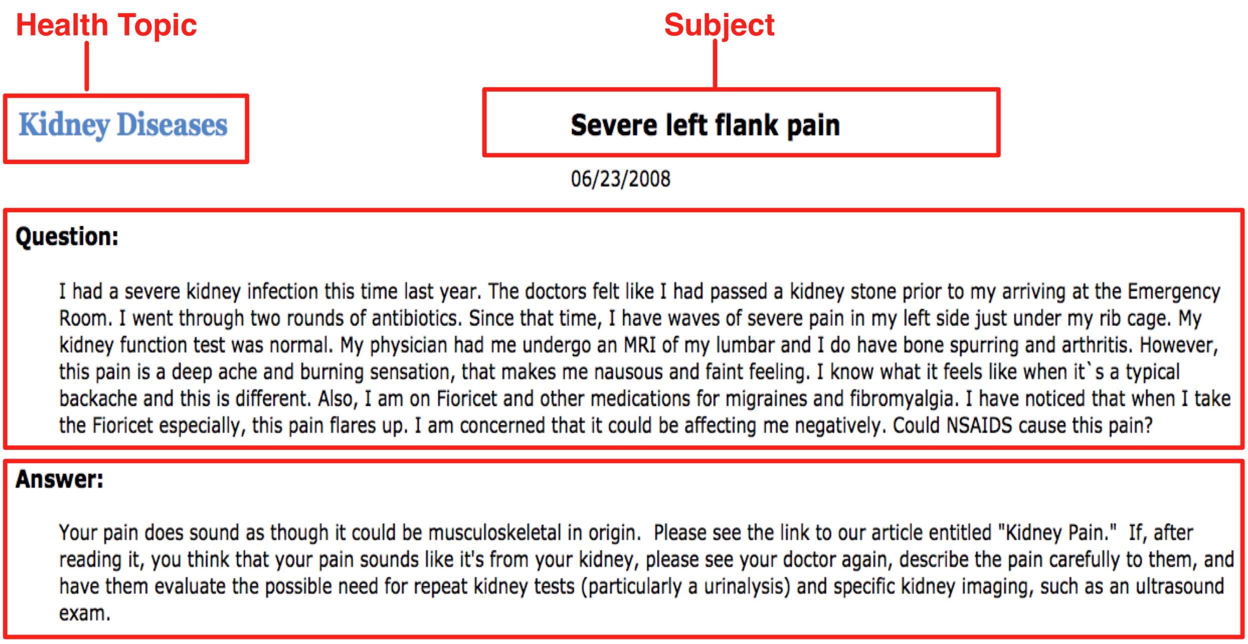
NetWellness is a non-profit Web service providing high-quality health information. It has been in operation since 1995 with over 13 million visits per year by consumers across the world in recent years. Consumer questions in NetWellness have been answered by medical and health professional faculties at three Ohio partner universities: Case Western Reserve University, the Ohio State University, and University of Cincinnati. However, each question was assigned a single topic, thereby limiting the potential benefit of using navigational exploration. For example, although the question in Figure 1 was assigned the topic "Kidney Disease", it can also be related to the topics "Pain Management" and "Pharmacy and Medications". Allowing for multiple relevant topics assigned to a single question (if applicable) enables consumers to reach it through multiple pathways, thus improving the retrieval recall in navigational exploration. To categorize health questions into multiple topics, in our previous study, we used MetaMap [19] to assign CUIs (Concept Unique Identifiers) in Unified Medical Language System [20] to these questions, since CUIs allow for the handling of synonyms. The CUI tags were also used for assigning questions to one or several of the 99 predefined topics, which took the semantics of the questions' contents into account.

In this study, we present a novel Conjunctive Exploratory Navigation Interface (CENI) for exploring NetWellness health questions with health topics as dynamic and searchable menus complementing lookup search. CENI provides a conjunctive mechanism for users to quickly drill down to relevant contents, rather than being exposed to an overwhelming number of webpages that are unlikely to be helpful.

To evaluate the effectiveness of CENI, we conducted a comparative study of search interfaces with anonymous, paid participants recruited from an online labor marketplace called Amazon Mechanical Turk (AMT) [21], a well-known and widely used crowdsourcing platform. AMT provides an attractive platform due to the relative ease of recruitment, low cost, and access to a diverse and large pool of potential participants. Crowdsourcing has been validated as a valuable method for conducting online experiments including health research [22,23], behavioral research [24,25], natural language processing [26-28], imaging analysis [29,30], drug discovery [31], and user interface evaluation [32]. Komarov et al investigated the validity of performing crowdsourcing evaluations of user interfaces, and the results provided evidence that AMT could be a productive mechanism for conducting performance evaluations of user interfaces to complement existing methodologies [32].

This study presents a novel conjunctive exploratory navigation mechanism to support consumer health information retrieval. Its efficacy is validated by conducting a crowdsourced comparative study of search interfaces for NetWellness consumer health questions.

Figure 1. A sample consumer health question in NetWellness, consisting of four components: Health Topic, Subject, Question, and Answer.



Kidney Diseases

Severe left flank pain

06/23/2008

Question:

I had a severe kidney infection this time last year. The doctors felt like I had passed a kidney stone prior to my arriving at the Emergency Room. I went through two rounds of antibiotics. Since that time, I have waves of severe pain in my left side just under my rib cage. My kidney function test was normal. My physician had me undergo an MRI of my lumbar and I do have bone spurring and arthritis. However, this pain is a deep ache and burning sensation, that makes me nausous and faint feeling. I know what it feels like when it`s a typical backache and this is different. Also, I am on Fioricet and other medications for migraines and fibromyalgia. I have noticed that when I take the Fioricet especially, this pain flares up. I am concerned that it could be affecting me negatively. Could NSAIDS cause this pain?

Answer:

Your pain does sound as though it could be musculoskeletal in origin. Please see the link to our article entitled "Kidney Pain." If, after reading it, you think that your pain sounds like it's from your kidney, please see your doctor again, describe the pain carefully to them, and have them evaluate the possible need for repeat kidney tests (particularly a urinalysis) and specific kidney imaging, such as an ultrasound exam.

Methods

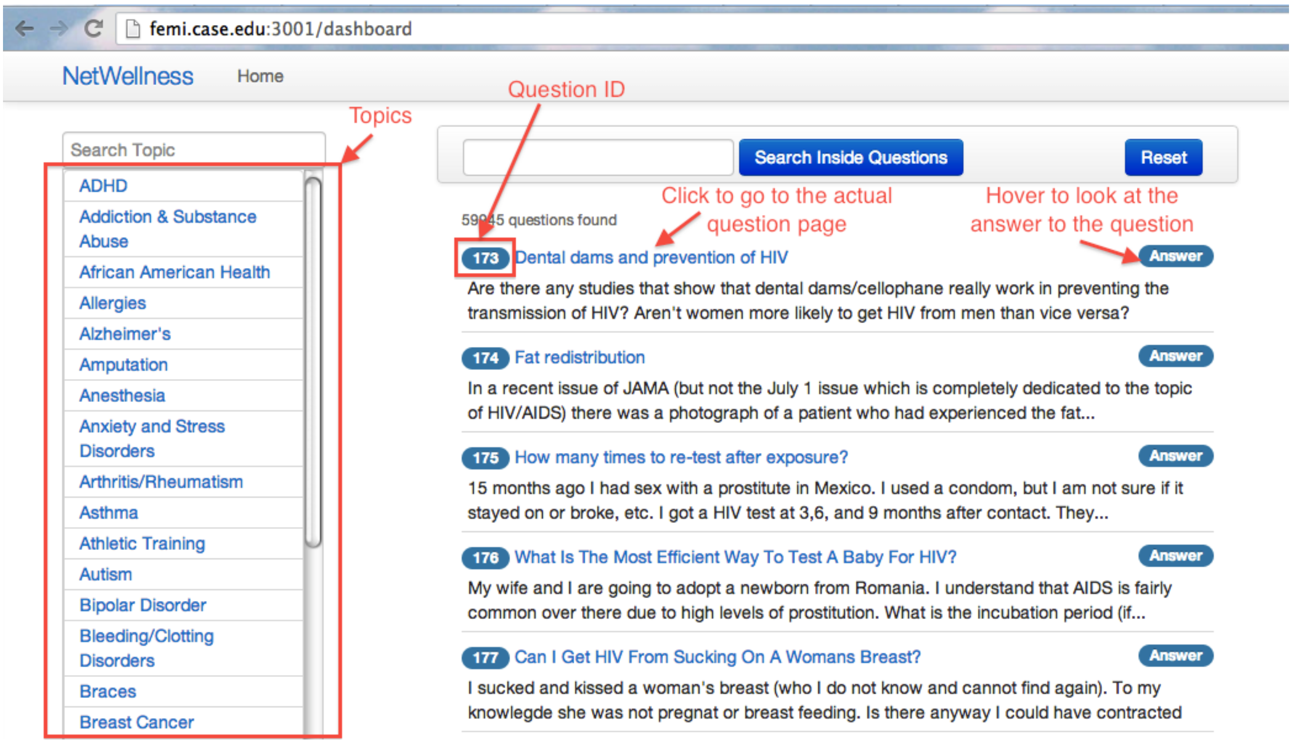
Data Corpus

The data corpus used in CENI consisted of over 60,000 consumer health questions in NetWellness dated through 2012. The CENI interface relies on the underlying organization of health questions by health topics, allowing for multiple relevant topics assigned to a single question so that users can reach the question through multiple pathways. A predefined set of 99 health topics were used for tagging each question with multiple topics in our previous study.

Conjunctive Exploratory Navigation Interface

We developed the CENI interface using agile Web development with Ruby on Rails [33]. Figure 2 is a screenshot of the CENI interface, where the left column displays a list of dynamic and searchable topic menus, and the right column contains health questions. By default, all the questions are displayed if neither topic nor search keyword is specified. Figure 3 shows a sample screenshot of CENI interface after the selection of topics “Depression” and “Pregnancy”, and the user specifying “anti-depressant” as a search term to search within the returned results. In this case, the right column displays the questions tagged with all the selected topics and containing the specified keyword. The chosen topics are displayed inside the horizontal bar on the top of the right column, where the “Reset” button is used to start a new exploration by clearing the specified topics and search terms. If only a single topic is selected, all the questions tagged with the topic will be displayed. Single topic selection is equivalent to the traditional navigational exploration using static menus.

Figure 2. The newly developed CENI interface: a topic-based navigational exploration interface.



59945 questions found

173 Dental dams and prevention of HIV

Are there any studies that show that dental dams/cellophane really work in preventing the transmission of HIV? Aren't women more likely to get HIV from men than vice versa?

174 Fat redistribution

In a recent issue of JAMA (but not the July 1 issue which is completely dedicated to the topic of HIV/AIDS) there was a photograph of a patient who had experienced the fat...

175 How many times to re-test after exposure?

15 months ago I had sex with a prostitute in Mexico. I used a condom, but I am not sure if it stayed on or broke, etc. I got a HIV test at 3,6, and 9 months after contact. They...

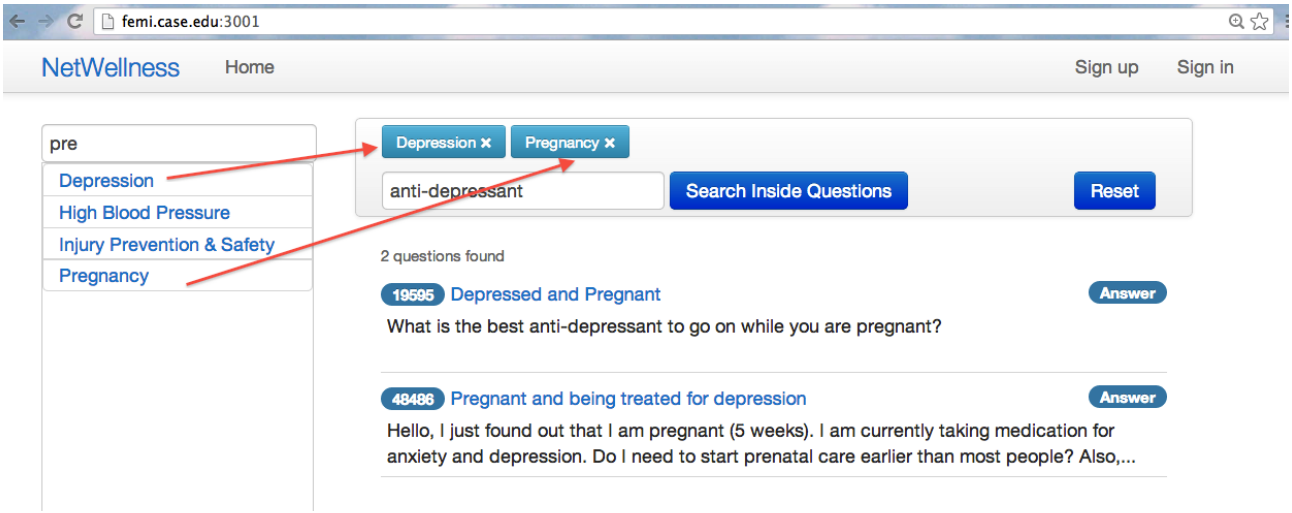
176 What Is The Most Efficient Way To Test A Baby For HIV?

My wife and I are going to adopt a newborn from Romania. I understand that AIDS is fairly common over there due to high levels of prostitution. What is the incubation period (if...

177 Can I Get HIV From Sucking On A Womans Breast?

I sucked and kissed a woman's breast (who I do not know and cannot find again). To my knowlegde she was not pregnat or breast feeding. Is there anyway I could have contracted

Figure 3. The CENI interface after selecting "Depression" and "Pregnancy" as topics, and specifying "anti-depressant" as keyword for search.



2 questions found

19595 Depressed and Pregnant

What is the best anti-depressant to go on while you are pregnant?

48486 Pregnant and being treated for depression

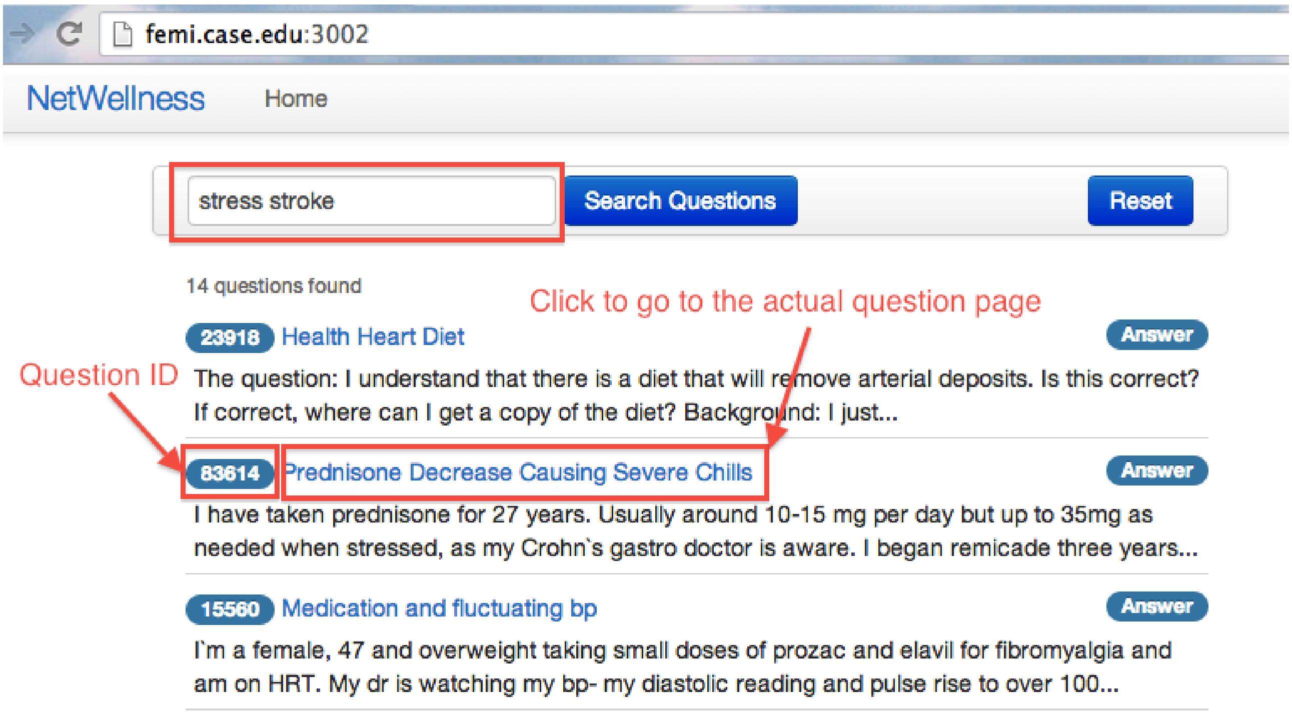
Hello, I just found out that I am pregnant (5 weeks). I am currently taking medication for anxiety and depression. Do I need to start prenatal care earlier than most people? Also,...

Crowdsourcing Comparative Evaluation of Search Interfaces

Search Interfaces Compared

To evaluate the efficacy of our newly developed CENI for navigating and exploring NetWellness health questions, we performed a comparative evaluation against two other search modes using keyword-based searches. One of them is an additional newly developed interface similar to the CENI interface but with topic menus removed (Figure 4). The other is the existing keyword-based search mode either using the advanced search interface provided by NetWellness official website [34] (Figure 5) or using Google search interface [35] with “NetWellness” as part of the keywords (Figure 6). In the following descriptions, we refer to “A” (Figure 2) as the newly developed CENI interface, “B” (Figure 4) as the additional newly developed search interface with keyword-based search only, and “C” as the existing keyword-based search mode either using the advanced search in NetWellness (Figure 5) or Google Search (Figure 6).

Figure 4. The newly developed keyword-based search interface.



stress stroke

14 questions found

23918 Health Heart Diet

The question: I understand that there is a diet that will remove arterial deposits. Is this correct? If correct, where can I get a copy of the diet? Background: I just...

83614 Prednisone Decrease Causing Severe Chills

I have taken prednisone for 27 years. Usually around 10-15 mg per day but up to 35mg as needed when stressed, as my Crohn`s gastro doctor is aware. I began remicade three years...

15560 Medication and fluctuating bp

I`m a female, 47 and overweight taking small doses of prozac and elavil for fibromyalgia and am on HRT. My dr is watching my bp- my diastolic reading and pulse rise to over 100...

Figure 5. The existing advanced search feature provided in NetWellness official website.

NetWellness is a global, community service providing quality, unbias from our partner university faculty. NetWellness is commercial-free advertising.

Search NetWellness

Search our collections:

☐ NetWellness HealthTopic articles

☐ NetWellness HealthTopic article Authors

☐ NetWellness Monthly Features, Newsletters, etc

☑ Ask an Expert questions and answers

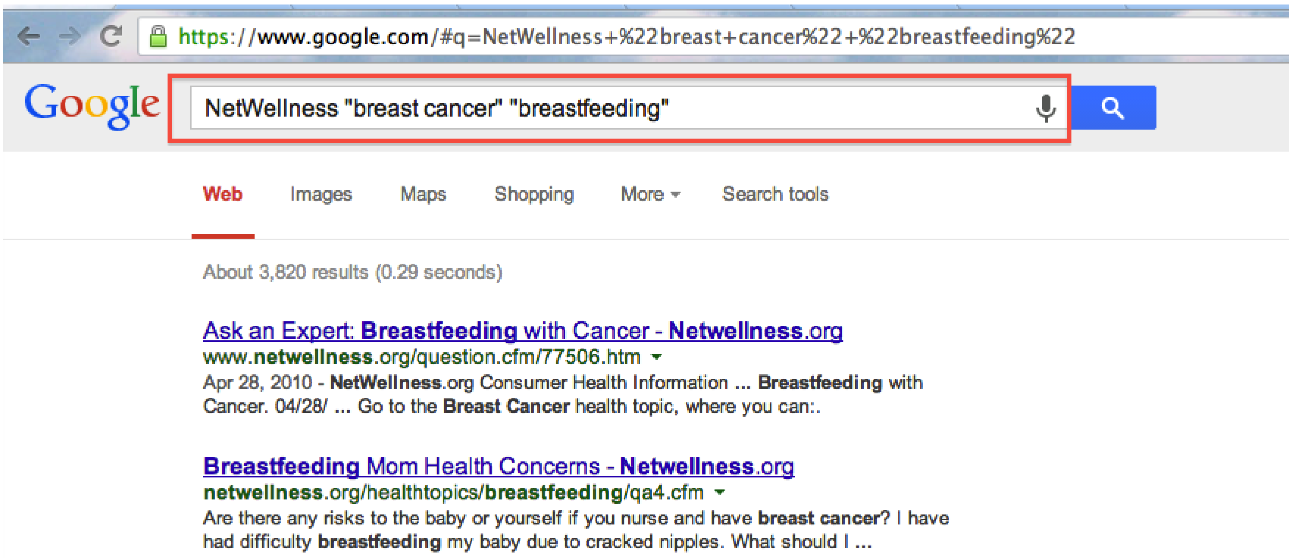
☐ Ask an Expert Forum Descriptions

☐ Ask an Expert Authors

☐ Weblinks

Advanced Search stress stroke

Figure 6. The existing Google Search interface with "NetWellness" as part of the keywords.



https://www.google.com/#q=NetWellness+%22breast+cancer%22+%22breastfeeding%22

NetWellness "breast cancer" "breastfeeding"

About 3,820 results (0.29 seconds)

Ask an Expert: **Breastfeeding** with Cancer - **Netwellness**.org
www.**netwellness**.org/question.cfm/77506.htm
Apr 28, 2010 - **NetWellness**.org Consumer Health Information ... **Breastfeeding** with Cancer. 04/28/ ... Go to the **Breast Cancer** health topic, where you can:.

Breastfeeding Mom Health Concerns - **Netwellness**.org
netwellness.org/healthtopics/**breastfeeding**/qa4.cfm
Are there any risks to the baby or yourself if you nurse and have **breast cancer**? I have had difficulty **breastfeeding** my baby due to cracked nipples. What should I ...

Search Tasks and Rationale

To perform a comparative study of the three search modes, we designed nine search tasks (Table 1) based on relevant health questions in NetWellness. Nine tasks were selected so that we could divide the tasks into three groups to be answered by each subject using three search modes in different orders, to avoid potential order effect, in a reasonable amount of time. Intuitively, the larger the number of tasks, the less biased the study design would be. However, using more than nine tasks would entail the need to allocate a greater amount of time to the AMT workers, making it harder to recruit them (although there were sufficient financial resources available to pay the workers).

Since lookup search is a well-established area, our search tasks were designed to be mostly exploratory in nature. Therefore, each search task involved at least two health topics and had one or more related health questions. Take the search task “What might be the concerns on breastfeeding while diagnosed with breast cancer?” as an example. It involved two topics “breastfeeding” and “breast cancer”. In contrast to questions with a clear answer using a unique search string, the nine tasks were deliberately selected to demonstrate multiple navigation pathways leading to them, representing areas where the distinct search mode could potentially be optimal.

Our study is focused on this question: “Are there tasks that may definitely benefit from conjunctive search?” This has not been established before because of a lack of an interface such as CENI to support conjunctive search. We selected the nine specific tasks based on the two information retrieval modes mentioned earlier (ie, lookup and exploratory search), and they are all based on the repository of over 60,000 consumer health questions collected. Three out of the nine tasks were “lookup” in nature, which are Tasks 3, 5, and 6. The other six tasks were designed to be “exploratory” in nature. They were selected based on several criteria: (1) they should be exploratory, (2) they should be relevant in health information seeking, and (3) they should preferably have multiple associated questions in the question repository that would provide the answer (otherwise standard search would likely be sufficient).

Table 1. List of the nine search tasks.

Task ID	Task description
1	What are the typical vision problems associated with diabetes?
2	What are possible relationships between Alzheimer’s disease and diet?
3	Can anti-epileptic medications be taken during pregnancy?
4	What are the possible connections between smoking and erectile dysfunction?
5	Can asthma be a side effect of taking Zocor?
6	Is colon cancer an inherited disease?
7	How might Tuberculosis medication impact one’s body weight?
8	Other than prescribed medications, what other approaches may help with depression?
9	What might be the concerns on breastfeeding while diagnosed with breast cancer?

Comparative Study Design and Procedure

We devised six surveys corresponding to six orders of search modes: ABC, ACB, BAC, BCA, CAB, CBA (see Multimedia Appendix 1 for the survey corresponding to the order ABC). To compare three search modes, we used the within-subjects design [36] with search mode as the factor, in three levels (A, B, C), which was counterbalanced appropriately to avoid order bias. In these six different surveys, the orders of the search tasks were kept the same while the orders of the search mode were permutated.

Each survey mainly consisted of three parts, where each part had three search tasks using one type of search mode. For each part, a brief instruction with screenshots (such as Figure 7) was provided to help participants better understand how the search task could be done with a particular search mode. Each search task included a question on the rating of the level of difficulty of the task, as well as three questions that served as validation of the quality of the search task that the participant worked on. For ratings, a 9-point Likert scale was used, where “9” indicated very difficult, “5” indicated neutral, and “1” indicated very easy. The rating responses were used as primary data for the comparative statistical evaluation of three search modes.

Because of the AMT setting, this study placed greater emphasis on obtaining valid answers by the workers. To ensure the validity of the participants’ responses, the remaining three questions required participants to provide the IDs of all the relevant health questions they found, the search keywords or topics they used, and a summary of the answers given from corresponding health questions. Each participant was required to provide answers to all the four questions unless he or she made five attempts without finding any results. In that case, the attempted search keywords or topics still needed to be recorded. Responses to these three questions were used to determine the quality of the AMT worker, and if the participant’s work would be accepted or rejected. Since our objective was to evaluate the effectiveness of topic-based navigational exploration interface A, for each search task using A, we required participants to provide at least two topics from the given list for each exploration they performed. If a worker provided answers that did not match the corresponding search mode, his or her work was rejected.

During the study design phase, we considered the alternative that all answers would be accepted, correct or not, and the rate of correctness would be an evaluation criterion. We decided against this because of the paid-nature of the AMT setting: If we did not insist in obtaining correct answers (which may not be unique), we could not be sure that AMT workers were not tempted to provide arbitrary answers; then we could not guarantee that they followed the instructions carefully. A worker not following the instructions and not required to get correct answers could potentially complete the tasks in a minimum amount of time by selecting arbitrary answers, and in effect earn much higher hourly pay. This potential conflict of financial interest from the worker would have been a weakness that might have rendered the study data less useful.

Each survey also included an additional set of background questions:

1. How frequently do you use Google search? (9-point Likert scale: 1-Always, 5-Occasionally, 9-Never)
2. How often do you search for health information online? (9-point Likert scale: 1-Always, 5-Occasionally, 9-Never)
3. How would you rate your level of medical knowledge? (9-point Likert scale: 1-None, 5-Average, 9-Expert)
4. Choose your level of education. (5-point Likert scale: 1-Less than high school, 2-High school, 3-College, 4-Graduate or Higher, 5-Other)
5. Among the three health information search approaches, which one do you prefer most and why to complete the above search tasks?

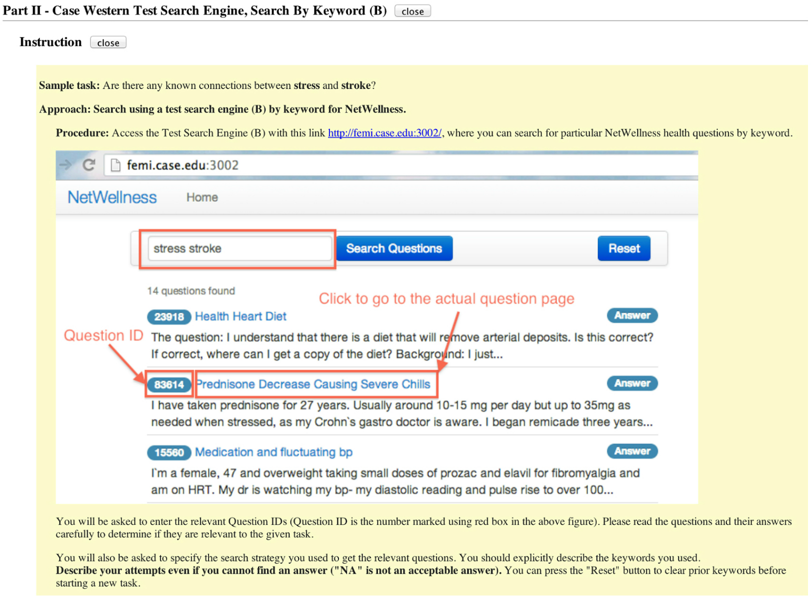
These additional questions were optional but could provide information for further analysis.

For each survey, we created a separate HIT (Human Intelligence Task, the unit of paid work) on AMT. Each HIT was allocated 60 minutes for the completion of the task with US $6 compensation; 15 participants were recruited per HIT. Each was required to complete nine tasks in Table 1 (three tasks for each search mode). All participants had an approval rate of at least 85%.

Before the six surveys were given to AMT workers, a pilot survey was given to a small group of 5 AMT workers to provide feedback on (1) time needed to complete the tasks, (2) clarity of the instructions, (3) whether the nine tasks made sense, and (4) other open comments on the study design. The pilot workers found 45 minutes to be reasonable for them but suspected that more time would be beneficial because of the variations in search experiences. There were no major comments for items (2), (3), and (4). Therefore, the actual surveys for the 90 AMT workers were allotted 60 minutes.

The study involved the use of survey and assessment procedures that were obtained in such a manner that the human subjects could not be identified directly or through identifiers linked to the subjects. Therefore the study qualified as an exempt research activity by the Case Western Reserve University Institutional Review Board under the Code of Federal Regulations, 38 CFR 16.101(b) Section 3, Category 2.

Figure 7. A brief instruction with screenshots for Part II.



Part II - Case Western Test Search Engine, Search By Keyword (B)

Instruction

Sample task: Are there any known connections between **stress** and **stroke**?

Approach: Search using a test search engine (B) by keyword for NetWellness.

Procedure: Access the Test Search Engine (B) with this link http://femi.case.edu:3002/, where you can search for particular NetWellness health questions by keyword.

femi.case.edu:3002

stress stroke

14 questions found

23918 Health Heart Diet

The question: I understand that there is a diet that will remove arterial deposits. Is this correct? If correct, where can I get a copy of the diet? Background: I just...

83614 Prednisone Decrease Causing Severe Chills

I have taken prednisone for 27 years. Usually around 10-15 mg per day but up to 35mg as needed when stressed, as my Crohn`s gastro doctor is aware. I began remicade three years...

15560 Medication and fluctuating bp

I`m a female, 47 and overweight taking small doses of prozac and elavil for fibromyalgia and am on HRT. My dr is watching my bp- my diastolic reading and pulse rise to over 100...

You will be asked to enter the relevant Question IDs (Question ID is the number marked using red box in the above figure). Please read the questions and their answers carefully to determine if they are relevant to the given task.

You will also be asked to specify the search strategy you used to get the relevant questions. You should explicitly describe the keywords you used. **Describe your attempts even if you cannot find an answer ("NA" is not an acceptable answer).** You can press the "Reset" button to clear prior keywords before starting a new task.

Results

Participant Responses

Data collection for crowdsourced participant responses took place from September 18 to October 6, 2013 (a period of 18 days); 98 eligible distinct AMT workers in total were recruited. Eight workers' responses were rejected because they were not able to complete the search tasks in the allotted time or did not follow the task instructions. As a result, the responses of 90 participants were used for the evaluation. These participants took an average of 50 minutes to complete a HIT, or survey.

Table 2 shows the numbers of participants who found relevant results using three search modes for the nine search tasks. Whether the participants found relevant results was manually validated based on their responses on the IDs of all the relevant health questions they found, the search keywords or topics they used, and a summary of the answers given in the corresponding health questions. Participants were allowed to give up a search task after five attempts were made without finding any relevant results. Tasks 3 and 8 received the least number of participants (82.2%, 74/90) who found relevant results. In sum, 90.9% (736/810) participants found relevant results: 96.3% (260/270) participants found relevant results using search mode A, 88.9% (240/270) participants found relevant results using B, and 87.4% (236/270) participants found relevant results using C.

For those who found no relevant results, manual review of the search keywords or topics provided showed that they did not devise appropriate keywords or topics. For example, five attempts made by one participant without success for Task 8 using search mode B were "depression treatments", "depression treatment without medication", "treating depression", "depression and meditation", and "depression and alternative treatments". The following five keywords were provided by another participant for Task 8 using search mode C: "depression", "no medication depression", "natural depression", "depression treatment", and "depression approach" with no relevant results found.

Table 2. Number of participants who found relevant results using three search modes (A, B, C) for the nine search tasks.

Task ID	Participants who found relevant results using A, n	Participants who found relevant results using B, n	Participants who found relevant results using C, n	Participants who found relevant results, n (%)
1	30	30	30	90 (100)
2	30	24	28	82 (91.1)
3	26	24	24	74 (82.2)
4	28	29	26	83 (92.2)
5	29	30	26	85 (94.4)
6	29	30	26	85 (94.4)
7	29	24	25	78 (86.7)
8	29	20	25	74 (82.2)
9	30	29	26	85 (94.4)
Total	260 (96.3%)	240 (88.9%)	236 (87.4%)	90.9%

The information on participant responses of the first four additional questions is presented in Table 3 (see also Multimedia Appendix 2), where the numbers of participants are reported only for points less than or equal to 5. Of the 90 participants, all reported the frequency of using Google search, 89 responded to the frequency of searching health information online, 89 reported their medical knowledge levels, and 68 reported their education levels. Furthermore, 36 used Google search frequently or more (40%), and 41 occasionally used Google search (45.6%). Regarding the frequency of searching health information online, 43 reported frequently or more (47.8%), and 29 reported occasionally (32.2%). For knowledge, 79 had limited or no medical knowledge (87.8%). For the education level, 45 reported "College" education (50%), and 13 reported "Graduate or Higher" (14.4%).

Participant responses on the preference of three search interface modes are presented in Multimedia Appendix 2. All 90 participants reported their preferences; 43 participants preferred A (47.8%), 23 favored B (25.6%), 22 preferred C (24.4%), one chose A or B (1.1%), and one chose A or C (1.1%). Multimedia Appendix 3 presents the explanations of participants for preferring certain search modes.

Table 3. Information on participant responses of the first four additional questions (including only points on a scale less than or equal to 5).

Question	Points (≤5)	Participants, n
Google search frequency		
	1-Always	5
	2-Very Frequently	9
	4-Frequently	22
	5-Occasionally	41
Health information search frequency		
	1-Always	3
	2-Very Frequently	8
	3	22
	4-Frequently	10
	5-Occasionally	29
Medical knowledge level		
	1-None	44
	2	32
	3-Limited	3
	4	7
	5-Average	3
Education level		
	2-High school	9
	3-College	45
	4-Graduate or Higher	13
	5-Other	1

Comparison of Three Search Modes

Participant responses for the difficulty ratings of search tasks using different search modes are also presented in Multimedia Appendix 2. For each participant, the ratings for A, B, and C were calculated by averaging the difficulty ratings of three search tasks corresponding to A, B, and C, respectively. The ratings data for each search interface mode was normally distributed.

We used the difficulty rating as a measure to compare the effectiveness of three search interface modes. Table 4 lists the average difficulty ratings of three search interface modes for the nine search tasks. For all the search tasks except Tasks 5 and 6, A was rated easier than both B and C. For all the search tasks except Tasks 2 and 8, B was rated easier than C. Figure 8 shows the mean difficulty ratings and error bars for the three search interface modes.

Table 4. Average difficulty ratings of three search modes (A, B, C) for the nine search tasks (1-very easy, 5-neutral, 9-very difficult).

Task ID	A	B	C
1	3.43	4	4.53
2	4.5	5.7	5.3
3	4.47	4.93	6.43
4	3.8	4.27	5.53
5	4.8	4.03	4.5
6	4.07	3.8	4.57
7	5	5.57	5.97
8	4.77	7	5.87
9	3.93	4.43	5.07
Average	4.31	4.86	5.31

Repeated-measures ANOVA analysis of the data showed statistically significant differences among the difficulty levels of A, B, and C ($P<.001$). Wilcoxon signed-rank test (one-tailed) between A and B showed that A was significantly easier than B ($P<.001$). Paired *t* tests (one-tailed) between A and C showed A was significantly easier than C ($P<.001$). Paired *t* tests (one-tailed) between B and C showed that B was significantly easier than C ($P=.014$). In sum, A (CENI) performed the best among the three.

We applied the Pearson correlation analysis to explore the effect of Google search frequency, online health information search frequency, medical knowledge, and education level on the difficulty ratings of the participants. No strong correlations were found.

According to participant comments on preferring a certain search mode most (see Multimedia Appendix 3), the most predominant reason they preferred A was that A was very easy to use and allowed users to quickly narrow down the relevant results, which is consistent with CENI’s design objective. The following are two examples of their reasons:

> *I liked being able to easily select multiple topics to narrow my search quickly. I found it to be the easiest search method because it decreased the number of unrelated search results.*
>
> *A was much, much easier, because the problem had already divided things into these categories and I did not have to fish around at what keywords or phrases would give me the results I was looking for. I could click on any number of general topics and type in something to narrow down the search and quickly get what I was searching for.*

Another reason frequently mentioned by participants was that they enjoyed the design element of the organizing information:

> *Explore and/or search using a test search engine by topics (A) for NetWellness. This was the easiest by far. You actually got what you wanted going this way and it was easier. I liked having the things on the side to use to begin the search.*
>
> *I found A to be the best approach. Having everything in a category made it easy to narrow things down fast.*

Some of the participants also noticed CENI’s benefit to less experienced users and those who are not sure exactly what to search:

> *Test Search Engine by Topic (A) is the most preferred by me, since it helps even a less experienced user to find and target to the topics and answers he is looking for. And for the experts, definitely, it helps to save a lot of time by optimizing the search by topics and using keywords to further filter the search.*
>
> *I liked the categories. It made it easier to find what I was looking for. I would think it would be very helpful when someone is not sure exactly what search terms to use. It also seems really helpful for someone who was just diagnosed, or has questions about a disease but aren't sure what they want to know.*

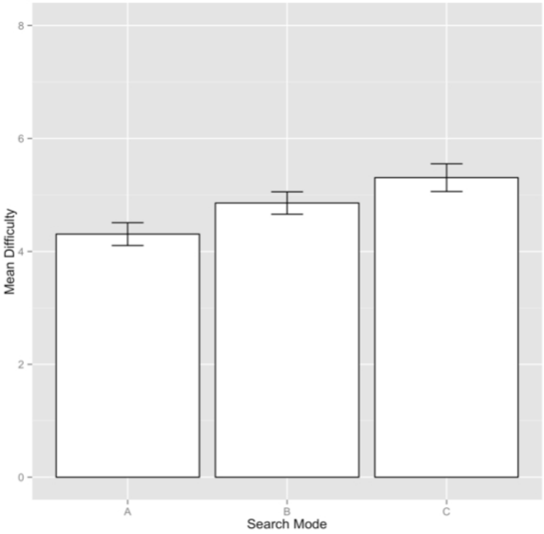
According to the participants who favored B (keyword-based search only), the most common reason was that it involved fewer steps to search: “I think approach B was easiest. So I prefer approach B. All I did was type a phrase I thought was relevant and it usually came up with relevant topics right away without further steps” and “I liked B the best because it was easier to use and less steps to search on it and it also gave good results. It was just clearer and less complicated than the others.”

We reviewed the keywords these participants (favoring B) provided for the search tasks, and it turned out that they were better at composing keyword phrases to acquire the most relevant results.

Among those who preferred C, the most common reason was that they were more familiar with it: “C, just because I’m more used to Google and also because I liked the way that Netwellness.org site was laid out” and “Netwellness.org/google. It was a familiar search engine to me and I found that it gave the quickest response to finding my answers.”

It is worth noting that for some participants, although they rated CENI (A) as the least difficult one, they still preferred other search modes (B or C) because they were more used to keyword-based search or Google search.

Figure 8. Mean difficulty ratings for search interface modes A, B, and C (1-very easy, 5-neutral, 9-very difficult) (error bar represents the standard error of the mean [SEM]; the SEMs for A, B, and C are 0.1015, 0.0999, and 0.1233, respectively).



Discussion

Principal Results

This study presented a novel navigational exploration interface CENI with topics as dynamic and searchable menus for consumer health information retrieval and navigation. Its conjunctive exploration mechanism allowed users to quickly drill down to the most relevant results. The efficacy of CENI was evaluated by comparing it with a similar search interface with keyword-based search only, as well as the existing search mode using Google search or NetWellness advanced search. The evaluation was conducted through crowdsourcing, a valuable method for gathering data when human participation is needed, which has been proved by many recent studies. To our knowledge, this work is the first crowdsourcing comparative evaluation of consumer health information retrieval and navigation interfaces.

We recruited 90 AMT workers as participants to work on six surveys (15 participants per survey). The surveys were carefully designed to include data quality control mechanisms and avoid order effect for comparison. The difficulty rating of search tasks using different search modes were used as a measure for the comparison. Statistical analysis showed significant differences among the difficulty rating of three search modes; CENI was significantly easier than the other two search modes. CENI was endorsed by 47.8% participants as the most preferred interface among the three search modes. Participants' reasons for preferring CENI most were consistent with its design objective and further validated its usability.

Although CENI was implemented for organizing and exploring consumer health questions in NetWellness, it is applicable to other domains for information organization and exploration as long as the information items can be classified into multiple categories.

Comparison With Prior Work

There were previous studies on health information retrieval and navigation [7,37]. Zeng et al [37] presented an interview and observation study in which consumers were asked to search for health information on MEDLINEplus [38,39], provided by National Library of Medicine. They analyzed the observed search sessions and determined several factors accounting for the failure of the specific searches, including confusing interface or organization of a website and information overload (too much information was retrieved). Our CENI interface presented a solution to address the potential information organization and overload challenges, by categorizing health information into one or more topics and using a conjunctive mechanism to quickly drill down to the most relevant contents.

Mu et al [7] provided a facet-view information navigation interface called SimMed for exploring literatures in a MEDLINE subset. They applied clustering technology to better organize users' exploratory search results. The effectiveness of SimMed was evaluated by comparing to a baseline system using knowledge changes, time spent, user-system interactions and patterns, and participant preference responses. Different from SimMed, CENI organized information items by topics or categories beforehand, instead of clustering them after a user's search. The effectiveness of CENI was measured using the quantitative difficulty level rated by the study participants and preference responses.

A unique aspect of our study is that we performed a crowdsourcing evaluation to compare search interfaces in a timely and cost-effective manner with a relatively large number of study participants recruited anonymously, rather than the traditional participatory recruitment in [7] and [37]. As far as we know, this is the first study of this kind performing direct comparison on consumer health information search interfaces.

Limitations of Using AMT

The first limitation of AMT is that the worker population has relatively higher levels of education (college or above) and is more tech-savvy compared to the general population, and therefore may not be representative [23,30]. However, large crowds like AMT workers are certainly more representative and cost-effective than the convenience samples in traditional participatory studies, which may consist of fewer participants due to time and labor constraints [7,30,40]. In addition, crowdsourcing provides access to a population that would not be readily available through traditional methods [23]. They are also more representative of the population that is familiar with and uses the Internet as an information source.

The second limitation for AMT is that it is harder to track time spent on individual tasks. AMT can keep track only of HIT time, and not subtasks within a HIT. Breaking down each search mode as a separate HIT would have the advantage of being able to automatically track time for each mode as a HIT level functionality provided by AMT. However, it would have two potential disadvantages: (1) it would not be a paired-design (which has more statistical power), since the same AMT worker would not have worked on all three modes, and (2) there would have been 18, as opposed to 6 HITs to manage, which would be substantial overhead for managing the study.

The third limitation of AMT is that it does not currently allow weblog information to be provided to track user-system interactions and patterns. We considered using weblogs for this purpose but did not pursue it in the end for two reasons: (1) it would have introduced an additional layer of complexity in an already complex survey (compared to typical AMT surveys), and (2) although we could have tracked IP addresses and weblog patterns for the site hosting CENI, obtaining weblogs from NetWellness’ official site and Google in order to identify the specific group of AMT workers who used their search facilities would be prohibitive due to the anonymous nature of AMT workers and the need to obtain data that are neither easily identifiable nor under our control.

The fourth limitation of AMT is a predefined time limit to complete the survey. An important factor enabling measuring user-system interactions and patterns mentioned in [7] is that there was no time limitation for participants to complete their search tasks. However, for each AMT HIT, a total allocated time must be specified beforehand.

A potential fifth limitation could be that an AMT worker might be biased towards the new interface to be evaluated in an effort to please the study designer and facilitate approval of their work. Our design implemented cross-validation questions and required intensity of focus to work on all nine search tasks, thereby minimizing this possibility. In fact, some evaluators still preferred C while in fact their rating for A is the easiest, indicating no intention to please the study designer.

Limitations of Study Design

The first limitation of our study was its use of the difficulty rating on individual tasks and modes as the major criterion. Tracking the amount of time an AMT worker took for each individual search mode would have been a useful source of information, as indicated in [7]. However, due to the limitation of AMT, it was less feasible to obtain this information without incurring other compromises in study design. We do suspect that less time would be correlated to lower levels of difficulty rating, however. And our overall preference rating and open commentary provided insight that this limitation did not affect the validity of the results.

The second potential limitation was that we selected nine search tasks, instead of a random selection of a larger number of tasks. Admittedly, more and randomly selected tasks would provide more statistical power. However, given the focused goal of the study to demonstrate the potential value of exploratory navigation using interfaces like CENI, the careful selection of search tasks was necessary, since our goal was not to show that CENI always performed better on all search tasks, lookup or exploratory.

The third potential limitation was that our CENI interface used the 99 existing NetWellness topics instead of developing an independent set of topics for tagging the questions. Our rationale was that a new set of topics dedicated for CENI would be an unfair advantage and would have introduced a major compounding factor, since the interface we compared against involved the exiting NetWellness resident search interface using the 99 topics. Also, we used the automatic multi-topic assignment results obtained in our previous study, where the precision and recall were 0.849 and 0.774 respectively. Nonetheless, this apparently did not affect our evaluation of the effectiveness of CENI, since it still outperformed the other two search modes. The correct and complete assignment of topics to questions would have definitely improved the effectiveness of the CENI interface more.

Conclusions

We presented a novel navigational exploration interface for consumer health information retrieval and navigation. With topics presented as dynamic and searchable menus, CENI’s conjunctive exploration mechanism allowed users to quickly drill down to the most relevant results in the most effective way.

AMT crowdsourcing allowed us to perform a comparative search interface evaluation in a timely and cost-effective manner with a relatively large number of study participants recruited anonymously. Our careful study, accounting for possible biases with cross-validation for results, has confirmed that CENI does enhance consumer health information access.

Acknowledgments

We thank Susan Wentz, Rong Xu, and Zhihui Luo for providing the consumer health information from NetWellness. This publication was made possible by the Clinical and Translational Science Collaborative of Cleveland, UL1TR000439 from the

National Center for Advancing Translational Sciences (NCATS) component of the US National Institutes of Health and NIH roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Authors' Contributions

LC and GQZ conceptualized and designed the CENI interface and the AMT evaluation surveys. RC provided valuable comments on the AMT setup and survey design. LC implemented the CENI interface with input from GQZ. LC and GQZ performed the AMT evaluation and data analysis. LC and GQZ led the writing of the paper, with critical reviews by RC.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Example of Web form used for performing evaluation of search interfaces in AMT.

[HTML File, 26KB - jmir_v16i2e45_app1.html]

Multimedia Appendix 2

Participant responses on the difficulty ratings of search tasks using different search modes.

[CSV File, 2KB - jmir_v16i2e45_app2.csv]

Multimedia Appendix 3

Participant explanations for preferring a certain search interface mode most.

[TXT File, 14KB - jmir_v16i2e45_app3.txt]

References

1. Fox S, Jones S. The social life of health information.: Pew Internet & American Life Project URL: http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information.aspx [accessed 2013-11-18] [WebCite Cache ID 6LEGVE27p]
2. Luo W, Najdawi M. Trust-building measures: a review of consumer health portals. Commun. ACM 2004 Jan 01;47(1):108-113. [doi: 10.1145/962081.962089]
3. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health Information on the Internet. JAMA 2001 May 23;285(20):2612-2621. [doi: 10.1001/jama.285.20.2612]
4. Broder A. A taxonomy of web search. SIGIR Forum 2002 Sep 01;36(2):3-10. [doi: 10.1145/792550.792552]
5. White RW, Kules B, Drucker SM, Schraefel M. Supporting Exploratory Search, Introduction. Commun. ACM 2006 Apr 01;49(4):36-39. [doi: 10.1145/1121949.1121978]
6. Hearst MA. Clustering versus faceted categories for information exploration. Commun. ACM 2006 Apr 01;49(4):59-61. [doi: 10.1145/1121949.1121983]
7. Mu X, Ryu H, Lu K. Supporting effective health and biomedical information retrieval and navigation: a novel facet view interface evaluation. J Biomed Inform 2011 Aug;44(4):576-586. [doi: 10.1016/j.jbi.2011.01.008] [Medline: 21277997]
8. Hearst MA. Design recommendations for hierarchical faceted search interfaces. 2006 Presented at: ACM SIGIR workshop on faceted search; Aug 10, 2006; Seattle, WA p. 1-5.
9. Dumais S, Cutrell E, Chen H. Optimizing search by showing results in context. 2001 Presented at: SIGCHI conference on Human factors in computing systems; March 31-April 5, 2001; Seattle, WA p. 277-284.
10. Kules B, Kustanowitz J, Shneiderman B. Categorizing web search results into meaningful and stable categories using fast-feature techniques. 2006 Presented at: 6th ACM/IEEE-CS Joint Conference on Digital Libraries; June 11-15, 2006; Chapel Hill, NC p. 210-219.
11. Kules B, Shneiderman B. Users can change their web search tactics: Design guidelines for categorized overviews. Information Processing & Management 2008 Mar;44(2):463-484. [doi: 10.1016/j.ipm.2007.07.014]
12. Wilson ML, Kules B, Shneiderman B. From Keyword Search to Exploration: Designing Future Search Interfaces for the Web. FNT in Web Science 2010;2(1):1-97. [doi: 10.1561/1800000003]
13. Stuckenschmidt H, Waard AD, Bhogal R, Fluit C, Kampman A, Buel JV, et al. A topic-based browser for large online resources. In: Engineering Knowledge in the Age of the Semantic Web. Berlin Heidelberg: Springer; 2004:433-448.
14. MEDLINE. URL: http://www.nlm.nih.gov/bsd/pmresources.html [accessed 2013-11-08] [WebCite Cache ID 6Kz6AuUTo]
15. WebMD Answers. URL: http://answers.webmd.com/ [accessed 2013-11-02] [WebCite Cache ID 6KpoYN3fA]

16. Health Category in Yahoo Answers. URL: http://answers.yahoo.com/dir/index?sid=396545018 [accessed 2013-11-02] [WebCite Cache ID 6Kpw9ZOht]
17. NetWellness. URL: http://netwellness.org/ [accessed 2013-11-02] [WebCite Cache ID 6Kpx1iMY9]
18. Morris TA, Guard JR, Marine SA, Schick L, Haag D, Tsipis G, et al. Approaching equity in consumer health information delivery: NetWellness. J Am Med Inform Assoc 1997;4(1):6-13 [FREE Full text] [Medline: 8988468]
19. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
20. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
21. Amazon Mechanical Turk. URL: https://www.mturk.com/mturk/ [accessed 2013-11-03] [WebCite Cache ID 6KqDlvlzc]
22. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. J Med Internet Res 2012;14(2):e46 [FREE Full text] [doi: 10.2196/jmir.1988] [Medline: 22397809]
23. Turner AM, Kirchhoff K, Capurro D. Using crowdsourcing technology for testing multilingual public health promotion materials. J Med Internet Res 2012;14(3):e79 [FREE Full text] [doi: 10.2196/jmir.2063] [Medline: 22664384]
24. Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods 2012 Mar;44(1):1-23. [doi: 10.3758/s13428-011-0124-6] [Medline: 21717266]
25. Crump MJ, McDonnell JV, Gureckis TM. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One 2013;8(3):e57410 [FREE Full text] [doi: 10.1371/journal.pone.0057410] [Medline: 23516406]
26. Snow R, O'Connor B, Jurafsky D, Ng AY. Cheap and fast - but is it good?: evaluating non-expert annotations for natural language tasks. 2008 Presented at: Conference on Empirical Methods in Natural Language Processing; October 25-27, 2008; Honolulu, Hawaii p. 254-263.
27. Saunders DR, Bex PJ, Woods RL. Crowdsourcing a normative natural language dataset: a comparison of Amazon Mechanical Turk and in-lab data collection. J Med Internet Res 2013;15(5):e100 [FREE Full text] [doi: 10.2196/jmir.2620] [Medline: 23689038]
28. Zhai H, Lingren T, Deleger L, Li Q, Kaiser M, Stoutenborough L, et al. Web 2.0-based crowdsourcing for high-quality gold standard development in clinical natural language processing. J Med Internet Res 2013;15(4):e73 [FREE Full text] [doi: 10.2196/jmir.2426] [Medline: 23548263]
29. Luengo-Oroz MA, Arranz A, Frean J. Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears. J Med Internet Res 2012;14(6):e167 [FREE Full text] [doi: 10.2196/jmir.2338] [Medline: 23196001]
30. Yu B, Willis M, Sun P, Wang J. Crowdsourcing participatory evaluation of medical pictograms using Amazon Mechanical Turk. J Med Internet Res 2013;15(6):e108 [FREE Full text] [doi: 10.2196/jmir.2513] [Medline: 23732572]
31. Lessl M, Bryans JS, Richards D, Asadullah K. Crowd sourcing in drug discovery. Nat Rev Drug Discov 2011 Apr;10(4):241-242. [doi: 10.1038/nrd3412] [Medline: 21455221]
32. Komarov S, Reinecke K, Gajos KZ. Crowdsourcing performance evaluations of user interfaces. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 02, 2013; Paris, France p. 207-216.
33. Ruby on Rails. URL: http://rubyonrails.org/ [accessed 2013-11-05] [WebCite Cache ID 6KuYAAXaU]
34. Advanced Search NetWellness. URL: http://www.netwellness.org/search/advancedsearch.cfm [accessed 2013-11-02] [WebCite Cache ID 6KqA1PYIB]
35. Google Search. URL: https://www.google.com/ [accessed 2013-11-02] [WebCite Cache ID 6Kq5fAgzH]
36. Martin DW. Doing psychology experiments. Australia: Thomson/Wadsworth; 2008.
37. 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]
38. MEDLINEplus. URL: http://www.nlm.nih.gov/medlineplus/ [accessed 2013-11-08] [WebCite Cache ID 6Kz1F7PVY]
39. Miller N, Lacroix EM, Backus JE. MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. Bull Med Libr Assoc 2000 Jan;88(1):11-17 [FREE Full text] [Medline: 10658959]
40. Kim H, Nakamura C, Zeng-Treitler Q. Assessment of pictographs developed through a participatory design process using an online survey tool. J Med Internet Res 2009;11(1):e5 [FREE Full text] [doi: 10.2196/jmir.1129] [Medline: 19275981]

Abbreviations

AMT: Amazon Mechanical Turk
CENI: Conjunctive Exploratory Navigation Interface
CUI: concept unique identifier
HIT: human intelligence task

Edited by G Eysenbach; submitted 20.11.13; peer-reviewed by Q Zeng, H Zhai, Y Zhang; comments to author 16.12.13; revised version received 10.01.14; accepted 28.01.14; published 10.02.14.

Please cite as:
Cui L, Carter R, Zhang GQ
Evaluation of a Novel Conjunctive Exploratory Navigation Interface for Consumer Health Information: A Crowdsourced Comparative Study
J Med Internet Res 2014;16(2):e45
URL: http://www.jmir.org/2014/2/e45/
doi:10.2196/jmir.3111
PMID:24513593

Viewpoint

Routine Real-Time Cost-Effectiveness Monitoring of a Web-Based Depression Intervention: A Risk-Sharing Proposal

Klemen Naveršnik[1], BSc; Aleš Mrhar[2], PhD

[1]Prototype Analytics, Sandoz Development Center, Lek Pharmaceuticals, Ljubljana, Slovenia

[2]Chair for Biopharmaceutics and Pharmacokinetics, Faculty of Pharmacy, University of Ljubljana, Ljubljana, Slovenia

Corresponding Author:
Klemen Naveršnik, BSc
Prototype Analytics
Sandoz Development Center
Lek Pharmaceuticals
SDC
Verovškova 57
Ljubljana, SI-1000
Slovenia
Phone: 386 1 5803459
Fax: 386 1 5681393
Email: klemen.naversnik@sandoz.com

Abstract

Background: A new health care technology must be cost-effective in order to be adopted. If evidence regarding cost-effectiveness is uncertain, then the decision maker faces two choices: (1) adopt the technology and run the risk that it is less effective in actual practice, or (2) reject the technology and risk that potential health is forgone. A new depression eHealth service was found to be cost-effective in a previously published study. The results, however, were unreliable because it was based on a pilot clinical trial. A conservative decision maker would normally require stronger evidence for the intervention to be implemented.

Objective: Our objective was to evaluate how to facilitate service implementation by shifting the burden of risk due to uncertainty to the service provider and ensure that the intervention remains cost-effective during routine use.

Methods: We propose a risk-sharing scheme, where the service cost depends on the actual effectiveness of the service in real-life setting. Routine efficacy data can be used as the input to the cost-effectiveness model, which employs a mapping function to translate a depression specific score into quality-adjusted life-years. The latter is the denominator in the cost-effectiveness ratio calculation, required by the health care decision maker. The output of the model is a “value graph”, showing intervention value as a function of its observed (future) efficacy, using the €30,000 per quality-adjusted life-year (QALY) threshold.

Results: We found that the eHealth service should improve the patient’s outcome by at least 11.9 points on the Beck Depression Inventory scale in order for the cost-effectiveness ratio to remain below the €30,000/QALY threshold. The value of a single point improvement was found to be between €200 and €700, depending on depression severity at treatment start. Value of the eHealth service, based on the current efficacy estimates, is €1900, which is significantly above its estimated cost (€200).

Conclusions: The eHealth depression service is particularly suited to routine monitoring, since data can be gathered through the Internet within the service communication channels. This enables real-time cost-effectiveness evaluation and allows a value-based price to be established. We propose a novel pricing scheme where the price is set to a point in the interval between cost and value, which provides an economic surplus to both the payer and the provider. Such a business model will assure that a portion of the surplus is retained by the payer and not completely appropriated by the private provider. If the eHealth service were to turn out less effective than originally anticipated, then the price would be lowered in order to achieve the cost-effectiveness threshold and this risk of financial loss would be borne by the provider.

(J Med Internet Res 2014;16(2):e67) doi:10.2196/jmir.2592

KEYWORDS

depression; medical economics; value-based purchasing

Introduction

Pricing and reimbursement decisions for new health care interventions are key to patient access to these treatments. Once an intervention is approved (based on evidence of its safety, efficacy, and quality), decision makers are faced with finding the value of the new treatment. Value statements are based on the objective of the health care system, which is normally maximization of health. Cost-effectiveness (CE) analysis is used to calculate the costs per unit of health (ie, quality-adjusted life-year [QALY]) for a given intervention. Generally, an intervention is considered cost-effective if its incremental cost-effectiveness ratio (ICER; unit: €/QALY) is below a predetermined threshold.

The level of CE threshold, adopted by the United Kingdom's National Institute for Health and Care Excellence (NICE) lies in the range of £20,000-£30,000 per QALY [1]. A recent study reduced the estimate to £18,317 per QALY [2]. Interventions with an ICER below this range are generally deemed cost-effective. Interventions exceeding this range do not present good value for money and would generally not be implemented, unless there is significant value shown in other domains (such as low budget impact or a treatment for a priority disease area). The existence of such an explicit cost-effectiveness threshold provides a clear and predictable signal of value to the private sector (ie, drug manufacturers) because it specifies what decision makers will regard as being cost-effective [3]. A value-based approach enables companies to reduce their prices to levels that assure that their products are cost-effective. On the other hand, companies may respond to such a policy by raising their prices well above the production costs if the perceived value of the product is still high [4]. In either case, the decision maker's goal is achieved, since health is purchased at a price below the cost-effectiveness threshold.

The reimbursement process has traditionally been binary: an intervention is either reimbursed or not. However, policies have emerged that expand the options, for example, linking coverage to evidence development [5]. Where data are insufficient to take an informed view on cost-effectiveness, then a risk-sharing approach could be adopted. This would require the company and payer to agree to a contract where the cost of the drug is reimbursed, contingent on claims of clinical effectiveness being realized in practice. If expected outcomes are not achieved, prices would need to be changed [6]. There is a surge of interest in risk-sharing schemes between health care payers and pharmaceutical companies in Europe. Performance-based risk sharing could produce efficient market equilibrium, achieved by adjustment of the price and post-launch evidence collection [7]. The effectiveness of the existing contracts, however, is limited, particularly due to high administration costs, lack of transparency, and conflicts of interest [8].

The intervention studied here is a novel eHealth service to support the treatment of patients with depressive symptoms, which aims at improving medication adherence and collaborative care management by combining Web-based and mobile-based systems. The eHealth service is applied in addition to the usual care over the Internet and mobile phones and allows active patient engagement and care management performed by trained psychologists and has been described previously [9]. Details on the intervention are included in Multimedia Appendix 1. The intervention was found to be very cost-effective in a previous study [10] with an ICER of €1430/QALY. There was, however, a high degree of uncertainty because efficacy data were based on a small pilot trial (46 patients) with high attrition. Varying assumptions regarding dropout pattern resulted in significantly lower cost-effectiveness [10]. Evidence regarding long-term benefits resulting from routine use is not available, and the risk that the same efficacy is not realized in real life would have to be borne by the payer if the treatment were to be reimbursed based on the existing data. Since social decision makers are not risk taking [11], we examine how this risk could be shared with (or shifted to) the service provider.

The purpose of our paper is to identify if and how an efficient risk-sharing scheme for a new depression treatment can be implemented. The objective of the agreement is to guarantee that the payer purchases health at a cost below the CE threshold. Our aim is to provide a clear definition of (1) what data need to be collected during routine use as well as how these data are used to calculate real-time cost-effectiveness, and (2) how the price is adjusted in order to meet the CE threshold.

Methods

Design

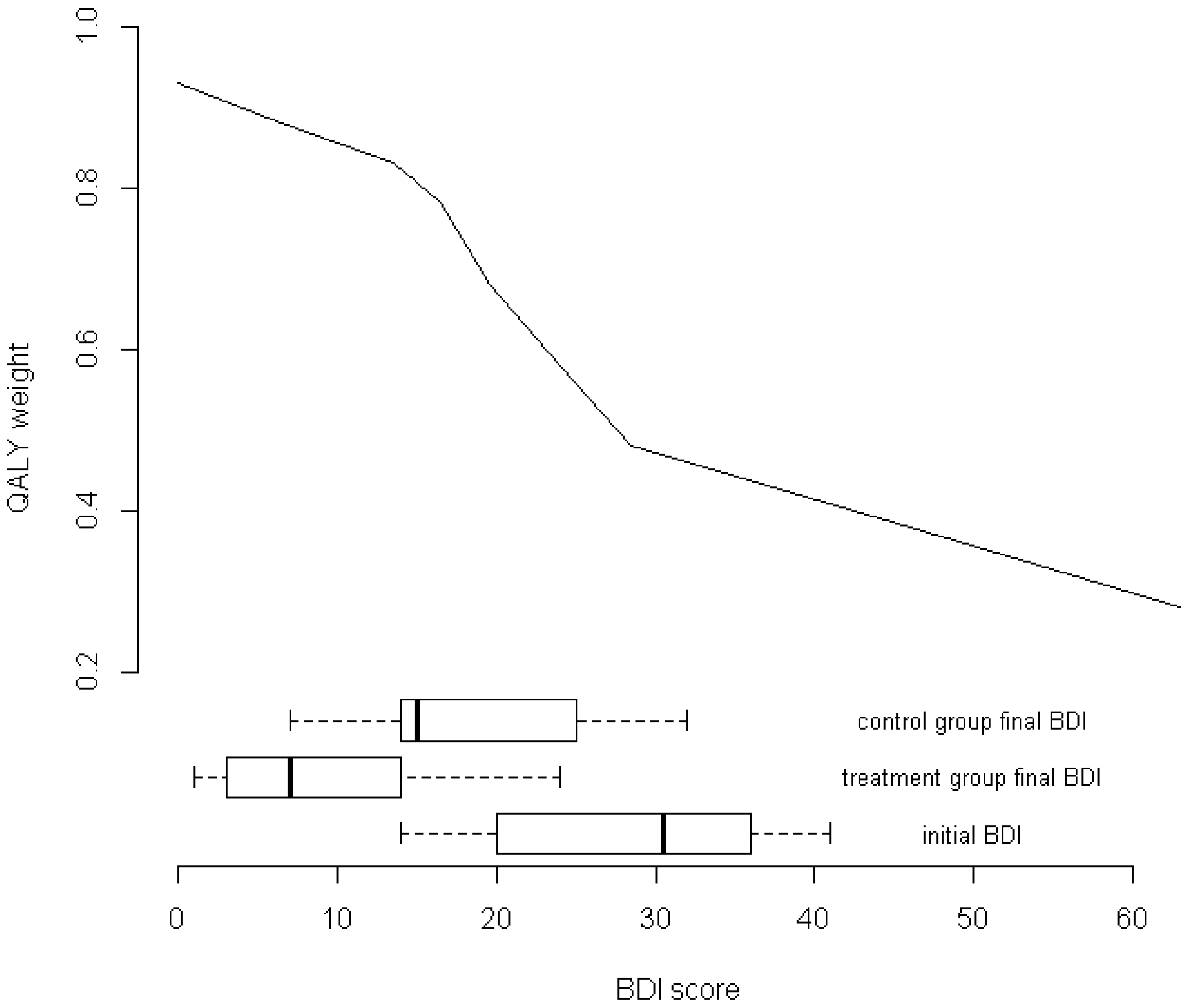
We propose a performance-based risk-sharing agreement to provide coverage "only in research" [12]. This means that the intervention is reimbursed only on condition that efficacy data are gathered for every patient during routine use. The objective is to periodically recalculate cost-effectiveness of the service based on the obtained data and adjust its price accordingly. If the recalculated ICER exceeds the CE threshold, the price is to be reduced. On the other hand, if the recalculated ICER is below threshold, the price may be increased and the surplus shared between the payer and provider.

Evidence to be Collected During Routine Use

Intervention efficacy is currently the main source of uncertainty. Long-term improvement of quality of life should thus be investigated during routine use. We propose that decline of depression severity be monitored by the Beck Depression Inventory (BDI-II) scale [13], self-administered by the patients. As evaluated in the pilot trial [9], the BDI score is collected at enrollment and at regular intervals during use. The questionnaire can be applied via the Health service platform and requires very few additional resources. Current mean estimates of BDI improvement, based on the pilot trial are shown as box-plots in Figure 1. The average BDI score for patients entering the trial was 29 ("initial BDI"). Half a year later, the score was reduced to 18 and 10 for the control and treatment arms respectively.

A previously published mapping function [10] is then used to convert BDI score into quality of life weight (Figure 1). Mapping of BDI scores onto the QALY scale is encouraged, since generic instruments such as EQ-5D lack sensitivity in measuring quality of life in mental disorders [14]. Collecting both EQ-5D and BDI data would be useful in order to validate the existing mapping function. Mapping BDI data onto QALY in the pilot trial (Figure 1) resulted in a mean initial QALY weight 0.53, which increased to final values 0.72 and 0.83 for the control and treatment arm respectively.

Figure 1. Box-plots of BDI scores for patients, entering the pilot trial (initial BDI) and treatment/control groups at the end of the pilot trial (piecewise linear line shows how BDI score [horizontal axis] is mapped onto the QALY scale [vertical axis]).



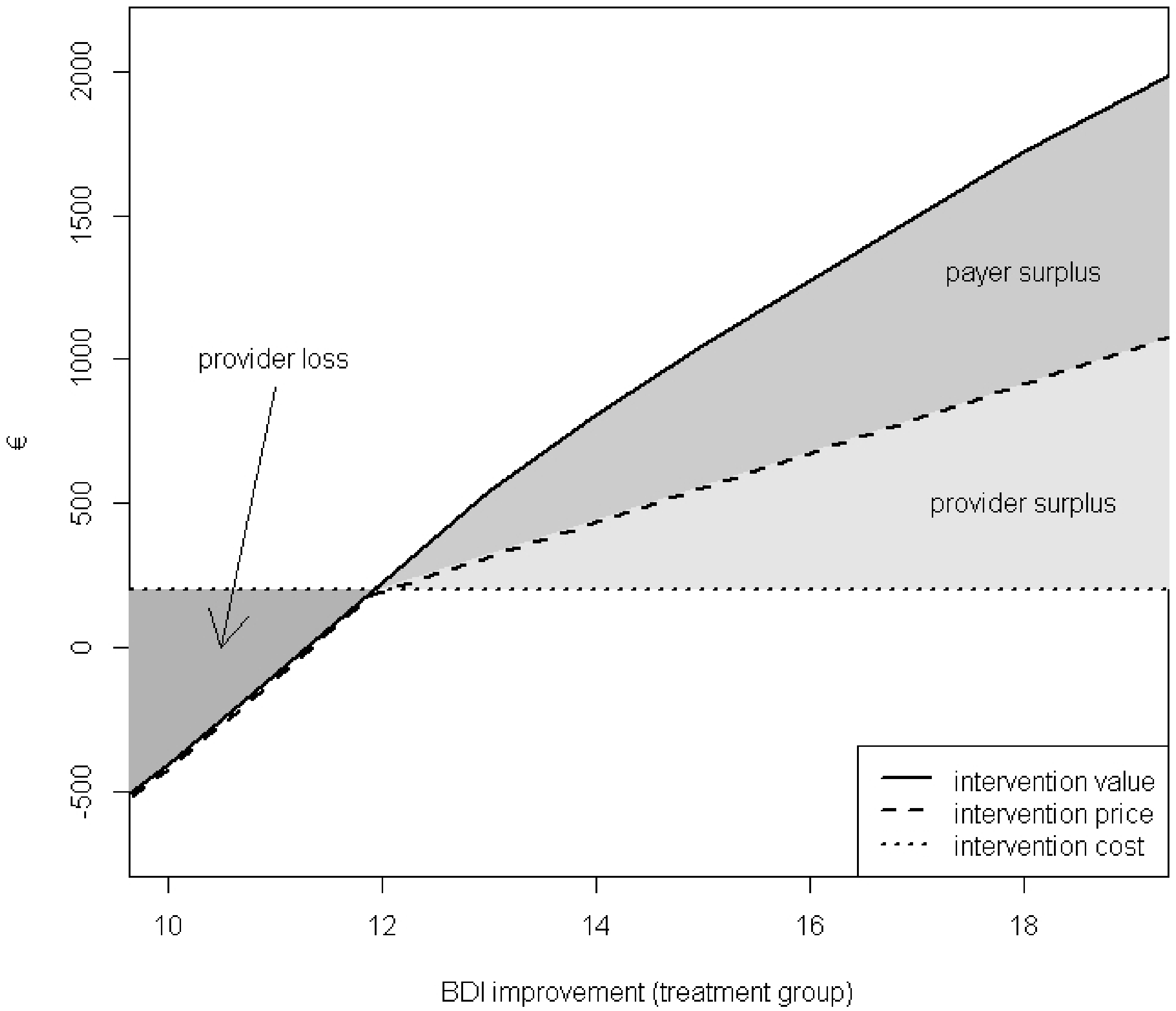
Results

Price Recalculation

How and when the evidence will be used to adjust the price should be clearly stated in the agreement. A previously published cost-effectiveness model [10] is used in order to calculate intervention value based on the data gathered along routine use. The economic model is detailed in Multimedia Appendix 1. Patients' BDI score improvement dictates the amount of QALYs gained (through the mapping function in Figure 1). A cost-effectiveness threshold of €30,000/QALY is used as an example in our calculations but would be adapted according to the payers' threshold. Explicitly stating the threshold monetary value for a QALY in turn allows calculation of the value of intervention. The price paid by the health authority should reflect the intervention's value, regardless of the cost borne by the provider (which is estimated at €200 per patient). Figure 2 shows how the price would be set, depending on the actual efficacy (BDI depression score reduction). The plot has two distinct regions, separated by the threshold BDI improvement level (11.9), which correspond to an ICER of €30,000/QALY (using the pilot trial BDI data, where control group BDI score improved by 10.9 points on average). In the low efficacy range (<11.9 BDI points), the provider is expected to reduce the price so that it does not exceed the intervention's value. Such a scenario would represent financial loss to the provider, since the price would have to be set below the actual intervention cost. On the other hand, if it turned out that efficacy is higher (>11.9 BDI points), intervention value would exceed its cost. The price could be raised in this case to a level not exceeding intervention value. Such a model would allow the total economic surplus (value minus cost) to be shared between the payer and provider. A portion of the surplus would go to the payer, since the intervention's value would exceed its price. The provider, on the other hand would retain the difference between the price charged and the intervention cost as profit. The actual service efficacy found in the pilot trial was a 19-point improvement on the BDI scale for the treatment arm. This translates into a value of €1900 (per patient per year), which is significantly above the estimated service cost.

The current market level of private appropriation in health technologies was estimated at about 15% [4]. An appropriate ratio of payer versus provider surplus would have to be agreed upon and would depend for instance on the perceived risk of provider loss and the amount of research and development investment in the creation of the service. Such an agreement would also prevent the service provider from gradually increasing the price until the cost-effectiveness is at (or just below) the threshold level.

Figure 2. Relation between intervention value, price, and cost, depending on the actual efficacy (BDI improvement) during routine use (intervention value calculated using €30,000/QALY threshold; shaded regions depict provider loss and social surplus, shared between the provider and payer at an example ratio of 1:1 [50% private appropriation level, intervention cost at €200 per patient]).



Data Collection

The key issue to be addressed is how to ensure a transparent mode of data collection. Practicality, small administrative burden, and low costs dictate that data be gathered through the provider's IT system. Means of ensuring data integrity would have to be set in place, as there is a clear conflict of interest if the provider is also the data collector. The data generation scheme would also need to provide control group data, since there is only a small historical control group available so far. This calls for a randomized design where an appropriate target control group size would need to be prespecified.

An important issue in analyzing the pilot trial data was how to handle a large body of missing data. Dropout rates up to 60% were observed [9] and should be reduced during routine use to reach a prespecified minimum response rate. Measures to reduce missing data could include a run-in period, limiting participants' burden in the data collection stage, and collecting data also for dropouts [15]. A major methodological issue to be addressed is how to handle missing data in the statistical analysis. The use of available cases for four alternative data imputation scenarios in the analysis of pilot trial data resulted in markedly different estimates of ICER. Handling missing BDI data should clearly be specified when calculating price adjustment, as well as a sensitivity analysis performed with alternative methods to check the robustness of results [16].

Discussion

Principal Findings

New drugs could have a substantial beneficial impact on service delivery and patient safety in practice, but it has been difficult to prove this within the confines of a phase III trial [8]. The situation is similar for a novel eHealth depression service, which appears to be very cost-effective yet has considerable uncertainty, since efficacy data are available only from a single pilot clinical trial. The decision maker must now consider whether the benefits of immediate access to the technology exceed the potential risk of the service being cost-ineffective in real practice. If the service is not implemented, a considerable amount of health may be forgone. On the other hand, if the service is implemented and turns out to be less cost-effective in actual practice, money invested could have been better spent on other health care interventions.

Due to its nature (applied over the Internet and mobile technologies), the eHealth depression service lends itself to routine data collection. In fact, many eHealth and mHealth interventions allow collection of effectiveness data during routine use, particularly if effectiveness can be self-reported by the patient, for instance by questionnaires. Such data can be very useful for an economic or any other evaluation, as it does not suffer from sampling limitations (ie, it effectively samples the whole treated population). This opens an array of options for market entry, such as risk-sharing agreements.

We propose that effectiveness data be gathered throughout routine use, once the service is launched. These data points, in turn, can be used to iteratively reassess the intervention’s cost-effectiveness. This allows a risk averse payer to be charged with a price no higher than the value it represents. There has been a lot of skepticism regarding risk-sharing agreements, propelled by the UK multiple sclerosis patient access scheme, which was perceived as a “costly failure” [17]. We have thus addressed components of a potential risk-sharing agreement, most likely to be critical to its effectiveness. Three key aspects were how to ensure transparency of data collection and evaluation, handling of missing data, and obtaining control group data.

The core component of the proposed scheme is price recalculation based on the observed effectiveness. This is where risk is shared between the provider (price goes down if effectiveness is low) and the payer (price increases if the observed value is high). The price ceiling is set with regards to the explicit cost-effectiveness threshold, set by the decision maker. This ensures that the service represents good value to the payer. Furthermore, we propose a novel business model of surplus sharing between the payer and the service provider. An appropriate provider/payer surplus ratio ensures that a portion of the profits due to high cost-effectiveness is appropriated by the payer (health insurance) and that the service provider still retains an incentive to optimize the service in order to achieve high effectiveness.

Limitations

The service provider would benefit from patient selection based on their initial BDI score. The value of a BDI point is not uniform across the whole BDI range (Figure 3) due to varying slope of the mapping function (Figure 1). This could give the provider an incentive to include patients only with initial BDI scores at around 25, since this would maximize treatment effect and value.

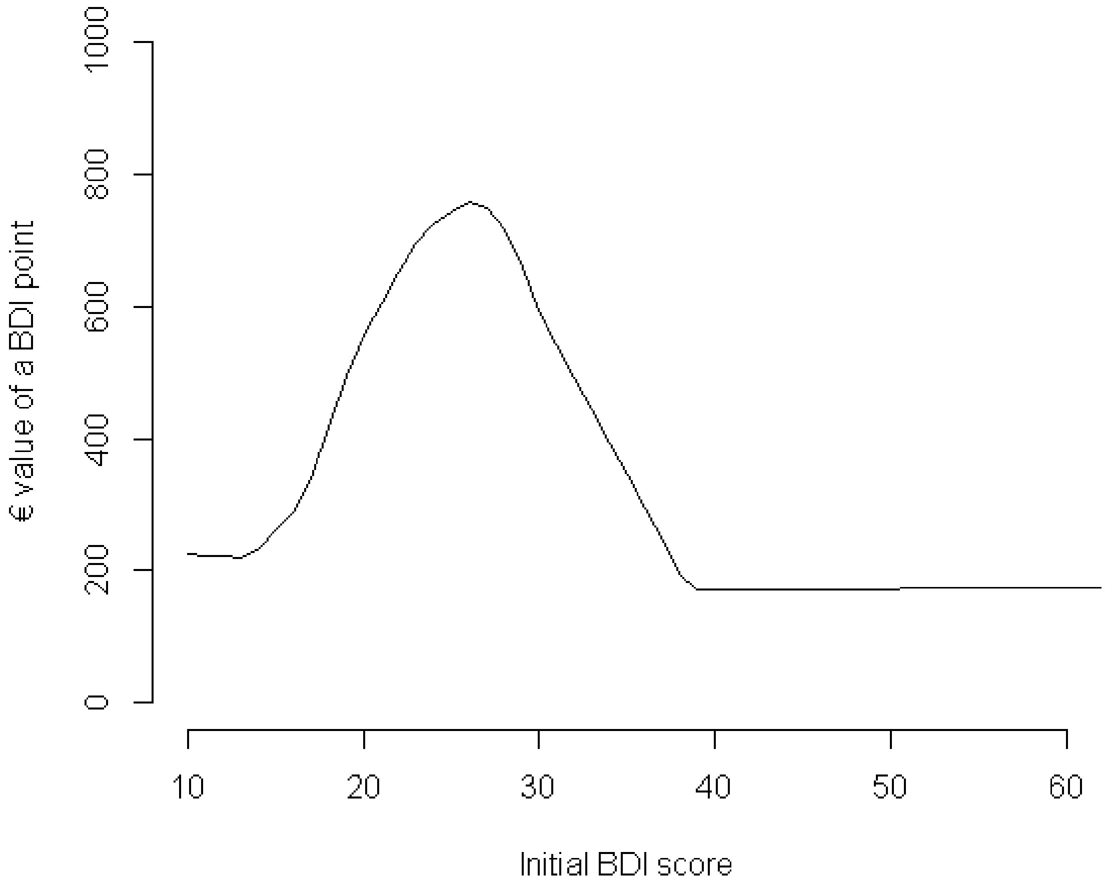
Another shortcoming is source of cost data used in the health economic model. Costs due to depression states were taken from a similar intervention in the United Kingdom, and these may be different in other settings. Although ICER was found to be insensitive to depression-related costs [10], these cannot be ignored, particularly because using the service could encourage patients to consume more health care resources (specialist visits, drugs, etc). Depression costs are significant in the indirect domain (productivity loss from absenteeism and presenteeism). These costs are not included in the model and would likely be reduced by treating depression. A conservative approach would thus assume that any potential cost increase due to higher health care consumption due to service would be offset by a reduction in indirect costs due to health improvement.

Voluntary consent to research, free of coercion, or penalty for refusal is a basic requirement for research involving competent adults [18]. When a treatment is offered in research only, this may have ethical implications [19]. If a patient was to decline reporting BDI data, would they still be eligible to use the eHealth service? If not, that may result in patient coercion to participate in the study. Random allocation to control and treatment groups could be an ethical issue if the service became standard treatment, since the service would be denied to patients, randomized to the control group. They would still receive treatment as usual but would be denied a potentially effective eHealth service. If this was deemed unacceptable, then a randomized control group would not be available. An alternative control group strategy is to follow up on patients that chose not to use the service (for instance due to lack of Internet access) and thus receive treatment as usual.

Our analysis is based on the notion that health (as measured by quality-adjusted life-years) is the only domain representing value to the payer. It should be noted that, depending on the perspective of the decision maker, other domains should also be taken into account, including severity of illness, unmet medical need, and wider societal considerations such as impacts on caregivers and equality [20]. Since value in other domains is difficult to monetize, we excluded them from the analysis.

Service cost is one of the variables in the price recalculation model. The current rough estimate of intervention cost is €200 per patient. The actual intervention cost (per patient) during routine use is likely to depend on the total number of patients enrolled in the program. Since cost is known only to the service provider, this could be exploited in order to achieve a higher price.

Figure 3. Value of a BDI point improvement (treatment minus control) as a function of the patient's initial BDI score (calculation based on a treatment effect of 10 BDI points).



Conclusion

We have shown how to shift the risk due to uncertainty from the health care payer to the service provider. Such an agreement requires a value-based pricing scheme and continuous efficacy monitoring throughout routine use. The latter is easily implemented in eHealth or mHealth interventions as long as a patient-reported measure of health is available. We proposed a flexible pricing scheme that allows economic surplus to be shared between the payer and provider in case of high treatment efficacy during routine use and allows the payer to avoid any potential loss in case of low efficacy.

Acknowledgments

No sources of funding were used to assist in the preparation of this manuscript. KN is a friend of one of the researchers in the pilot trial who has an interest in economic exploitation of eHealth services. All calculations were performed in *R: A language and environment for statistical computing*, and the code can be obtained from the authors. The authors thank both reviewers for their useful comments.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Intervention pilot study details.

[PDF File (Adobe PDF File), 41KB - jmir_v16i2e67_app1.pdf]

References

1. NICE. 2008. Guide to the Methods of Technology Appraisal URL: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf [accessed 2013-02-27] [WebCite Cache ID 6EklXJMJU]
2. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. CHE Research Paper 81. York, UK: Centre for Health Economics, University of York; 2013. Methods for the Estimation of the NICE Cost Effectiveness Threshold URL: http:/

/www.york.ac.uk/media/che/documents/papers/researchpapers/CHERP81_Methods_estimation_NICE_costeffectiveness_threshold.pdf [accessed 2013-02-27] [WebCite Cache ID 6EklQRM98]
3. Claxton K. OFT, VBP: QED? Health Econ 2007 Jun;16(6):545-558. [doi: 10.1002/hec.1249] [Medline: 17535009]
4. Jena AB, Philipson TJ. Cost-effectiveness analysis and innovation. J Health Econ 2008 Sep;27(5):1224-1236. [doi: 10.1016/j.jhealeco.2008.05.010] [Medline: 18619695]
5. Walker S, Sculpher M, Claxton K, Palmer S. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. Value Health 2012 May;15(3):570-579. [doi: 10.1016/j.jval.2011.12.013] [Medline: 22583469]
6. Office of Fair Trading. OFT: The Pharmaceutical Price Regulation Scheme: an OFT market study. 2007. URL: http://www.oft.gov.uk/shared_oft/reports/comp_policy/oft885.pdf [accessed 2013-02-27] [WebCite Cache ID 6EklNvwju]
7. Towse A, Garrison LP. Can't get no satisfaction? Will pay for performance help?: toward an economic framework for understanding performance-based risk-sharing agreements for innovative medical products. Pharmacoeconomics 2010;28(2):93-102. [doi: 10.2165/11314080-000000000-00000] [Medline: 20085386]
8. Adamski J, Godman B, Ofierska-Sujkowska G, Osińska B, Herholz H, Wendykowska K, et al. Risk sharing arrangements for pharmaceuticals: potential considerations and recommendations for European payers. BMC Health Serv Res 2010;10:153 [FREE Full text] [doi: 10.1186/1472-6963-10-153] [Medline: 20529296]
9. Meglic M, Furlan M, Kuzmanic M, Kozel D, Baraga D, Kuhar I, et al. Feasibility of an eHealth service to support collaborative depression care: results of a pilot study. J Med Internet Res 2010;12(5):e63 [FREE Full text] [doi: 10.2196/jmir.1510] [Medline: 21172765]
10. Naveršnik K, Mrhar A. Cost-effectiveness of a novel e-health depression service. Telemed J E Health 2013 Feb;19(2):110-116. [doi: 10.1089/tmj.2012.0081] [Medline: 23374036]
11. Zivin JG. Cost-effectiveness analysis with risk aversion. Health Econ 2001 Sep;10(6):499-508. [Medline: 11550291]
12. Trueman P, Grainger DL, Downs KE. Coverage with Evidence Development: applications and issues. Int J Technol Assess Health Care 2010 Jan;26(1):79-85. [doi: 10.1017/S0266462309990882] [Medline: 20059784]
13. Beck AT, Steer RA, Brown G. BDI-II, Beck depression inventory: manual. San Antonio, TX: Psychological Corporation; 1996.
14. Brazier J. Is the EQ-5D fit for purpose in mental health? Br J Psychiatry 2010 Nov;197(5):348-349 [FREE Full text] [doi: 10.1192/bjp.bp.110.082453] [Medline: 21037210]
15. Li T, Hutfless S, Dickerson K, Scharftstein D, Neaton J, Hogan J, et al. Minimal Standards in the Prevention and Handling of Missing Data in Observational and Experimental Patient Centered Outcomes Research. 2012. URL: http://www.pcori.org/assets/Minimal-Standards-in-the-Prevention-and-Handling-of-Missing-Data-in-Observational-and-Experimental-Patient-Centered-Outcomes-Research.pdf [accessed 2013-02-27] [WebCite Cache ID 6EklIKF4G]
16. Marshall A, Billingham LJ, Bryan S. Can we afford to ignore missing data in cost-effectiveness analyses? Eur J Health Econ 2009 Feb;10(1):1-3. [doi: 10.1007/s10198-008-0129-y] [Medline: 18936993]
17. Raftery J. Multiple sclerosis risk sharing scheme: a costly failure. BMJ 2010;340:c1672. [Medline: 20522654]
18. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? JAMA 2000;283(20):2701-2711. [Medline: 10819955]
19. Miller FG, Pearson SD. Coverage with evidence development: ethical issues and policy implications. Med Care 2008 Jul;46(7):746-751. [doi: 10.1097/MLR.0b013e3181789453] [Medline: 18580395]
20. Sussex J, Towse A, Devlin N. Operationalizing value-based pricing of medicines : a taxonomy of approaches. Pharmacoeconomics 2013 Jan;31(1):1-10. [doi: 10.1007/s40273-012-0001-x] [Medline: 23329588]

Abbreviations

BDI: Beck Depression Inventory
CE: cost-effectiveness
ICER: incremental cost-effectiveness ratio
QALY: quality-adjusted life-years

Edited by G Eysenbach; submitted 27.02.13; peer-reviewed by S Langrial, J Huber; comments to author 24.04.13; revised version received 30.05.13; accepted 16.01.14; published 27.02.14.

Please cite as:
Naveršnik K, Mrhar A
Routine Real-Time Cost-Effectiveness Monitoring of a Web-Based Depression Intervention: A Risk-Sharing Proposal
J Med Internet Res 2014;16(2):e67
URL: http://www.jmir.org/2014/2/e67/
doi:10.2196/jmir.2592
PMID:24583773

Review

Social Media: A Review and Tutorial of Applications in Medicine and Health Care

Francisco Jose Grajales III[1], BHK (Hons), MSc, CSEP-CEP; Samuel Sheps[2], MD, MSc, FRCPC; Kendall Ho[3], MD, FRCPC; Helen Novak-Lauscher[3], PhD; Gunther Eysenbach[4], MD MPH, FACMI

[1]eHealth Strategy Office, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[2]School of Population and Public Health, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[3]eHealth Strategy Office, Faculty of Medicine, University of British Columbia, Port Moody, BC, Canada

[4]Centre for Global eHealth Innovation, Techna Institute and Institute Health Policy, Management and Evaluation, University Health Network and University of Toronto, Toronto, ON, Canada

Corresponding Author:
Francisco Jose Grajales III, BHK (Hons), MSc, CSEP-CEP
eHealth Strategy Office
Faculty of Medicine
University of British Columbia
855 West 10th Avenue
Vancouver, BC, V5Z 1L7
Canada
Phone: 1 7788814061
Fax: 1 7788814061
Email: cisco@franciscograjales.com

Abstract

Background: Social media are dynamic and interactive computer-mediated communication tools that have high penetration rates in the general population in high-income and middle-income countries. However, in medicine and health care, a large number of stakeholders (eg, clinicians, administrators, professional colleges, academic institutions, ministries of health, among others) are unaware of social media's relevance, potential applications in their day-to-day activities, as well as the inherent risks and how these may be attenuated and mitigated.

Objective: We conducted a narrative review with the aim to present case studies that illustrate how, where, and why social media are being used in the medical and health care sectors.

Methods: Using a critical-interpretivist framework, we used qualitative methods to synthesize the impact and illustrate, explain, and provide contextual knowledge of the applications and potential implementations of social media in medicine and health care. Both traditional (eg, peer-reviewed) and nontraditional (eg, policies, case studies, and social media content) sources were used, in addition to an environmental scan (using Google and Bing Web searches) of resources.

Results: We reviewed, evaluated, and synthesized 76 articles, 44 websites, and 11 policies/reports. Results and case studies are presented according to 10 different categories of social media: (1) blogs (eg, WordPress), (2) microblogs (eg, Twitter), (3) social networking sites (eg, Facebook), (4) professional networking sites (eg, LinkedIn, Sermo), (5) thematic networking sites (eg, 23andMe), (6) wikis (eg, Wikipedia), (7) mashups (eg, HealthMap), (8) collaborative filtering sites (eg, Digg), (9) media sharing sites (eg, YouTube, Slideshare), and others (eg, SecondLife). Four recommendations are provided and explained for stakeholders wishing to engage with social media while attenuating risk: (1) maintain professionalism at all times, (2) be authentic, have fun, and do not be afraid, (3) ask for help, and (4) focus, grab attention, and engage.

Conclusions: The role of social media in the medical and health care sectors is far reaching, and many questions in terms of governance, ethics, professionalism, privacy, confidentiality, and information quality remain unanswered. By following the guidelines presented, professionals have a starting point to engage with social media in a safe and ethical manner. Future research will be required to understand the synergies between social media and evidence-based practice, as well as develop institutional policies that benefit patients, clinicians, public health practitioners, and industry alike.

(J Med Internet Res 2014;16(2):e13) doi:10.2196/jmir.2912

KEYWORDS

social media; blogging; social network

Introduction

Background

Social media are Web-based tools that are used for computer-mediated communication. In health care, they have been used to maintain or improve peer-to-peer and clinician-to-patient communication, promote institutional branding, and improve the speed of interaction between and across different health care stakeholders. Examples of social media applications in health include (but are not limited to) access to educational resources by clinicians and patients [1-3], generation of content rich reference resources (eg, Wikipedia) [4], evaluation and reporting of real-time flu trends [5], catalyzing outreach during (public) health campaigns [6,7], and recruitment of patients to online studies and in clinical trials [8-11].

A number of indicators suggest that the evidence for using social media in the health care context is growing; for example, the number of articles indexed on PubMed has nearly doubled each year for the last 4 years [12], social media policies are being adopted [13] and tested in various health care settings [14], journals are discussing how social media facilitate knowledge-sharing and collaboration [15,16], and theories on the social changes resulting from their adoption are being developed [17]. However, despite these useful insights, our collective understanding of how social media can be used in medical and health care remains fragmented.

Objective

The aim of this narrative review was to gain a better understanding of how social media are being used in health care. Using a qualitative approach, this article uses case studies to illustrate where, how, and why social media are being used. The intent of this review is to allow different health care stakeholders the opportunity to make informed decisions on how to use social media and similar electronic-mediated communication tools as part of their daily activities.

Methods

Qualitative Method

Although literature reviews in medicine have traditionally followed positivistic epistemologies, we drew upon a different approach, the critical-interpretivist theory [18], to conduct this review. Our intent, more specifically, was to elucidate impact while illustrating, explaining, and providing the contextual knowledge of why social media are being used in medicine and health care. However, we did not intend to measure, quantify, or generalize results, as is the case with Cochrane Reviews. Ultimately, the knowledge synthesized herein will allow readers to decide for themselves where, how, and why they may use and implement these computer-mediated communication tools as part of their day-to-day activities.

Data Sources and Knowledge Synthesis

This review used a number of traditional and nontraditional reference sources. It is not exhaustive due to inherent limitations that occur when trying to assess the medical and health-related grey literature situated within social media itself. Medline was searched using the search string in Textbox 1. Additionally, data from the Cochrane list of Web 2.0 resources [19], the Health Librarianship Canada (HLCanada) wiki [20], the Pan-American Health Organization’s Equity and Human Development Listserv, the 2008-2011 proceedings of the Medicine 2.0 World Congress on Social Media and Web 2.0 in Health, Medicine and Biomedical Research [21], and award winning blogs (eg, ScienceRoll) were used to supplement peer-reviewed resources. Where necessary, results were further supplemented with an environmental scan of the Google and Bing search engines.

Results were categorized based on social media service type (see definitions below) and, where necessary, further subgrouped as appropriate. The search above was conducted on January 1, 2012; however, all systematic reviews published after this date and up to July 2013 that fit these keywords were added to the literature.

Textbox 1. MEDLINE search string (modified from [12]).

("second life" AND (virtual OR 3d OR immersive)) OR "virtual worlds" OR "web 3.0" OR "medicine 2.0" OR "health 2.0" OR "web 2.0" OR mashup OR "social media" OR Blog OR digg OR "del.icio.us" OR "social bookmarking" OR wikis OR folksonomy OR wikipedia OR flickr OR twitter OR youtube OR facebook OR myspace OR Linkedin OR FourSquare

Definitions

Table 1 [22-26] presents a series of definitions and examples of different social media services.

Table 1. Categorical definitions of social media.

Service type	Definition	Example
Blog	Short for “web log”: a blog is an easy-to-publish website where bloggers (authors of blogs) post information and essays in sequential order [22].	WordPress, Blogger
Microblog	A tiny blog service that allows networks of users to send short updates to each other in less than 140 characters. Microblogs are considered a platform for information dissemination, social networking, and real-time communication [22].	Twitter, Identi
Social networking site	A social networking site is an online service, platform, or site that focuses on building and visualizing social networks or social relations among people, who, for example, share interests and/or activities. A social network service essentially consists of a representation of each user (often a profile), their social links, and a variety of additional services [23].	Facebook, MySpace
Professional networking site	A professional networking site is a type of social network service that is focused solely on interactions and relationships related to business or a person’s professional career [24].	LinkedIn, Sermo, Asklepios, Ozmosis, Drs Hangout, Doc2Doc
Thematic networking sites	Social networking sites centered on a particular theme; for example disaster response, nursing, etc. These share many aspects of, and operate as a community of practice.	Telehelp, Innocentive, 23andMe, PatientsLikeMeCureTogether
Wiki	Wikis are used to denote communal websites where content can be quickly and easily edited. Wikis support collaboration and information sharing; feature multimedia, such as video, slides, photographs; and allow anyone to edit or are password protected [22].	Wikipedia, Fluwiki
Mashups	A website that combines data and functionality from two or more services to create a new, value-added, service [25].	HealthMap, Google FluTrends
Collaborative filtering sites	A website where information is filtered or collected according to patterns. Techniques involving collaboration among multiple agents, viewpoints, and data sources are often used. These agents engage through a variety of sites, through a process called crowdsourcing, where the crowds join forces for a common purpose [26].	Digg, Delicious
Media sharing sites	A hosting service that allows individuals to upload and create galleries of photos, videos, and other digital media (eg, slide presentations). The host will then store them on a server and make them either publicly or privately available.	SlideShare, YouTube, Flickr
Other	Multi-User Virtual Environments, also known as Virtual Worlds	Second Life

Results

Summary

A total of 76 articles, 44 websites, and 11 policies/reports were reviewed and synthesized. Ethics, professionalism, privacy, and confidentiality, as well as information quality were recurrent themes throughout the literature. These are synthesized throughout the manuscript and emphasized in the discussion section. Prospectively, our results are presented based on the definitions of social media categories presented in Table 1.

Blogs

The first (Web 2.0) social media were developed in the late 1990s in the form of Web-logs (a term which was later shortened to blog). Web-based software platforms like Open Diary enhanced accessibility of content by allowing any existing or new users in the lay public to create a communal website where opinions about any topic could be voiced to create communal, collaborative dialogues. Blogs foster open access to information (both opinions and facts), contribute largely to the number of new websites created on the Internet and are often picked up by mainstream media, which makes them an important vehicle for social change [27]. For example, Paul Levy, the former President and CEO of Beth Israel Deaconess Centre in Boston, MA, was one of the early adopters among health care executives who, as a public authority with significant power, wrote a blog to make his reflections and decisions at the hospital transparent to all [28].

Although the literature on the use of blogs in Medline is growing, only one study that formally evaluated the attitudes, perceptions, and realities of the medical blogosphere was identified. Kovic and colleagues [27] conducted an online survey of medical bloggers and found that successful medical bloggers are most often highly educated writers (with masters or doctoral-level degrees) who are faithful to their sources and readers and are motivated to influence how others think by sharing their practical knowledge or skills in a creative manner.

Educational institutions in health care (eg, The Mayo Clinic) have used blogs to foster reflective peer-to-peer learning, which allow for open discussions and a formal log of medical training, as well as the implementation of new protocols [29]. Many hospitals also use blogs for branding and community outreach to (1) advertise their facilities [28] (such as featuring the newest device or test that they competitors do not possess), (2) share positive patient experiences [30] (such as a Mayo Clinic's Piano Foyer Video where 2 unrelated patients meet and play the piano together while in hospital), or (3) feature well-known physicians who treat famous people or athletes [31]. Hospitals also use blogs to disseminate disease-specific information to supplement leaflets or handouts for patient education [32].

Blogs have been used in clinical research for clinical trial recruitment and data collection, allowing patients to ask questions about the trial procedures, risks, and incentives while maintaining an anonymous, non-threatening environment [33]. Mayo Clinic has also used blogs focused on major depressive disorder to request feedback on the patient experience and some of the complementary medicine practices they follow [1].

Patients have also been using blogs creatively to monitor and share their own patient journeys. For example, SixUntilMe (named after the age at which the author, Kerri Morrone Sparling, was diagnosed with diabetes) features the life of a patient living with type 1 diabetes, discussing topics like insulin pumps, continuous glucose monitors, and diabetes advocacy [34]. Cancer patients have also used blogs to share their experiences with chemotherapy. Dave deBronkart, a well-known e-patient advocate, used his blog to inform family members and his attending (family) physician of changes in tumor growth from a self-created spreadsheet of radiology reports of tumor size data [35]. Also, in May 2011, the Vancouverite Derek K Miller had a friend post his auto-obituary after dying from stage 4 metastatic colorectal cancer. His self-obituary blog post [36] "went viral", receiving more than 4 million views in the 4 days after his death—a rather startling example of the potential reach of health-related social media.

Blogs have been used by health care workers for peer-to-peer communication and knowledge exchange such as virtual rounds. The Clinical Cases Blog [37] is prototypical of the medical blogosphere, as it features cases in allergy and immunology, cardiology, pulmonology, gastroenterology, nephrology, endocrinology, hematology, rheumatology, infectious diseases, neurology, geriatrics, and pain management. Moreover, this blog also has a special section on admission note templates (eg, congestive heart failure), procedure guides, and related material. Figure 1 displays an example case from this blog.

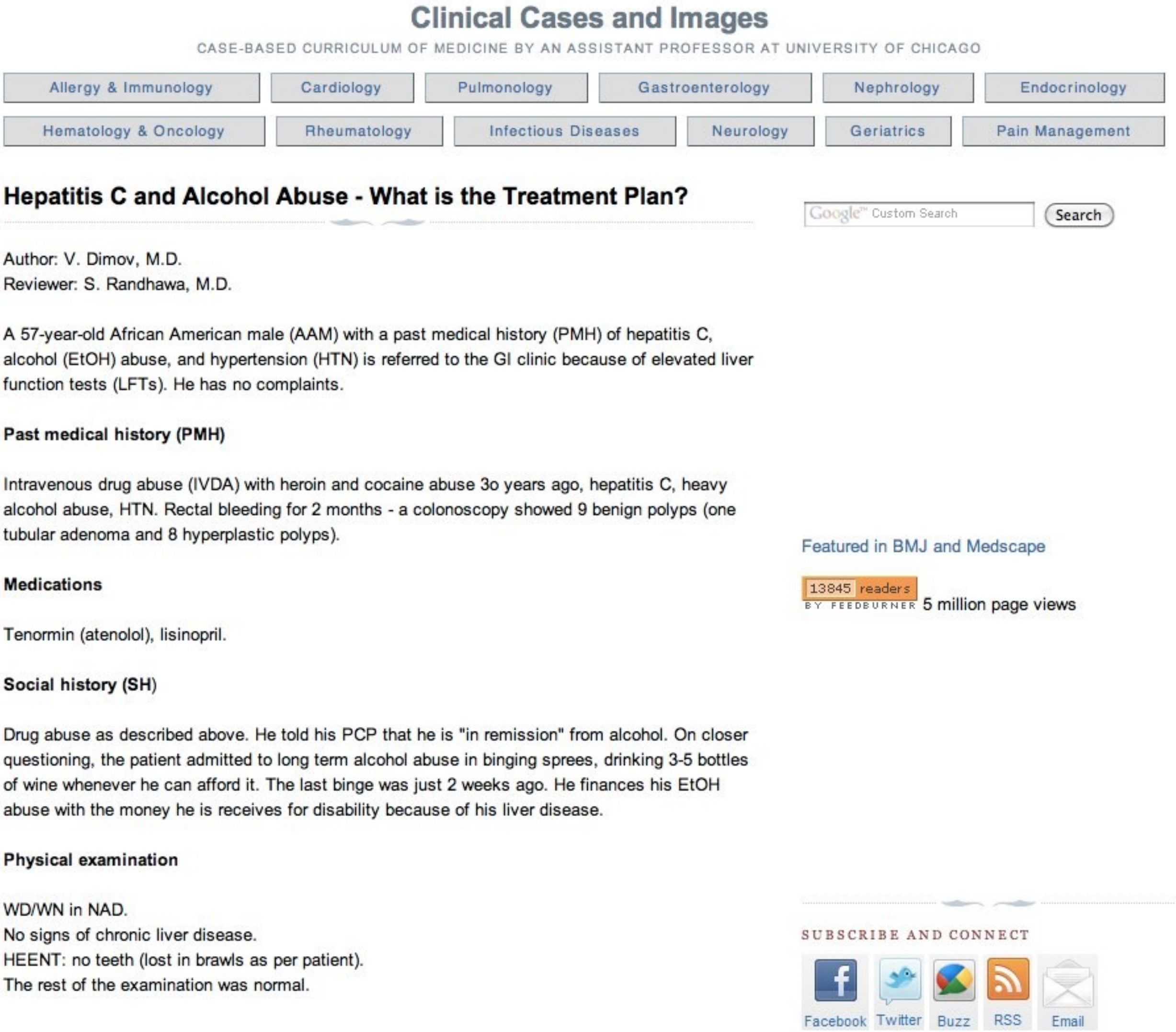
With regard to disease and epidemic outbreak tracking, citizen-report photo blogs have been used to inform hospitals of incoming mass casualty events (eg, Hudson River plane landing) [38]. Equally interesting, is how the US military has used natural language processing (where computers evaluate meaning) to automatically filter and retrieve information on blog posts by military servicemen as a means to monitor emotions and posttraumatic stress disorder after operational deployment [39].

There are many other examples of medical blogs in addition to the ones discussed. MedGadget [40] showcased a range of interests in 2010 with their annual (winter) medical blog competition using a public voting system. To see the winners, further illustrating how blogs are used in health care, see Multimedia Appendix 1.

The Really Simple Syndication (RSS) Web standard has facilitated the broad adoption and dissemination of blogs. RSS allows software, known as RSS readers, to pull content and create an email-like inbox of blogs and other websites (eg, PubMed) that are frequently updated. This is useful when a user wants to create a customized "feed" of information that is relevant to their interests and classify it accordingly, for easy retrieval in the future. Among the most notable RSS readers are Apple Mail, iGoogle, and Bloglines.

Overall, blogs are the oldest, most established, and evaluated form of social media, with articles as early as 2004 noting their use in medicine and family practice [41]. A number of peer-review articles on blogs have also been published. These mainly note their effectiveness how they can be used to disseminate best practices [42], their applications in assessing clinical knowledge learned [43], and how they can be used to promote reflection and professional development [29].

Figure 1. A sample rounding blogging case.



Hepatitis C and Alcohol Abuse - What is the Treatment Plan?

Author: V. Dimov, M.D.
Reviewer: S. Randhawa, M.D.

A 57-year-old African American male (AAM) with a past medical history (PMH) of hepatitis C, alcohol (EtOH) abuse, and hypertension (HTN) is referred to the GI clinic because of elevated liver function tests (LFTs). He has no complaints.

Past medical history (PMH)

Intravenous drug abuse (IVDA) with heroin and cocaine abuse 3o years ago, hepatitis C, heavy alcohol abuse, HTN. Rectal bleeding for 2 months - a colonoscopy showed 9 benign polyps (one tubular adenoma and 8 hyperplastic polyps).

Medications

Tenormin (atenolol), lisinopril.

Social history (SH)

Drug abuse as described above. He told his PCP that he is "in remission" from alcohol. On closer questioning, the patient admitted to long term alcohol abuse in binging sprees, drinking 3-5 bottles of wine whenever he can afford it. The last binge was just 2 weeks ago. He finances his EtOH abuse with the money he is receives for disability because of his liver disease.

Physical examination

WD/WN in NAD.
No signs of chronic liver disease.
HEENT: no teeth (lost in brawls as per patient).
The rest of the examination was normal.

Microblogs

The most dynamic and concise form of information exchange on social media occurs on microblogs. These short 21st century telegrams allow users to view a large number of updates, of brief content, over a short period of time. Today, a large number of microblogging platforms exist, catering to audiences varying from the corporate world to teenagers; however, Twitter is and has remained the most prominent service on the market. Twitter updates are known as tweets.

Newcomers to Twitter often perceive the character restriction as a barrier to communication; however, this misconception usually decreases with repeated use, as tweets are easily supplemented with shortened hyperlinks to other digital media, such as videos or websites. Historically, a 140-character limit was chosen to allow interoperability with SMS (short message service) text messages.

“Tweeps” (people who tweet) also often use other services that connect to the Twitter platform (eg, TweetDeck, HootSuite), which allow them to organize their tweets, manage information, and see website previews or pictures without having to click on a link and open a new Web browser window. Some of these services also sort, filter, and curate tweets, allowing a user to see updates related to a particular topic (eg, health care issues). In turn, this has caused a new technical tweeting language to emerge. A summary glossary of this language can be found in Multimedia Appendix 2, adapted from the Twitter Glossary.

Tweets and tweeting styles can be classified in three broad categories [44]. Substantive tweets are independently understandable (eg, a tweet with an abridged title or author of a paper, a brief comment, and a link to the publication, or a headline teaser to a blog). Conversational tweets are fragments of a new or ongoing conversation that draw on professional or personal interests or comment on current events. Finally, there are hybrid tweets, which are substantive and conversational at the same time (eg, “discussing my supervisor’s newest Nature publication at the Mahoney and Sons pub”).

In medicine and health care, there have been over 140 reported uses for Twitter [45]. There have been some interesting applications of Twitter in medical education. The Pennsylvania State College of Medicine has used Twitter to augment peer-to-peer and instructor-to-student learning [46] by stimulating topic discussions, providing feedback on critical thinking, conducting course evaluations, disseminating writing prompts, soliciting class responses, and monitoring student progress. Second, a junior doctor and a medical student started a Twitter Journal Club [47] that functions in the same manner as traditional journal clubs, except that the means for discussion is Twitter. By using a combination of blog posts, where the paper and discussion questions are posted in advance, along with the hashtag #TwitJC, students, doctors, and anyone interested in the subject can engage and interact in a meaningful way. The club meets every second Sunday evening.

The use of Twitter at conferences can also be seen as a medical education application [48]. In this case, Twitter functions as a tool to discuss and enhance the speaker's presentation in real-time through the comments of the audience. The Medicine 2.0 conference series has pioneered the use of Twitter-screens, displaying the tweets of the audience alongside the presenter slides.

In terms of health service delivery, 2 physicians have used @tweetspreekuur since October 2009 for primary care consultations [49]. Using the concept of learning by doing, the service was launched with little planning. After 1 year, their tweeting practice has shown that consultations encompass all areas of primary care, though the main reasons for contact are advice, reassurance, and triage. Typically, Tweet exchanges vary from one to eight tweets in length and about one third of the communication takes place publicly, while the other two-thirds takes place through direct messages (which are not public). Pictures of skin and genital-related problems have also been sent to the service. Presently, @tweetspreekuur is run on a voluntary basis (there is no reimbursement to the physicians who run the service) and the "attending physicians" involved stipulate that their success is due to language (consultations take place in Dutch, limiting their audience) and the option for patients to continue the consultation through a secure online platform, only available in the Netherlands. Today, preliminary research [50] suggests that Twitter has been effective at providing access to care at a low cost, that running the service is fun and entertaining for the providers, and that the level of user satisfaction is high.

In this section, it is important to draw attention to hashtags, which are a form of information curation that allow people to find tweets related to a particular discussion or topic. Among the most common are #HCSM or Health Care Social Media and its Canadian (#hcsmca), European (#HCSMEU), and Latin American (#HCSMLA) variants, #Med2, #MDChat, and #Health20. For example, a tweet that has both the #Med2 and #HCSMCA hashtags will be read by people who filter tweets because of their interest in information related to social media in Canada, as well as the Medicine 2.0 conference series.

On the negative side, there are a large number of bots (short for "robots") that re-tweet and spam Twitter users who use particular words, phrases, or hashtags in order to increase their user reach and digital footprint. Fortunately, these phony users are typically removed by the Twitter service relatively quickly due to the "report spam" feature on the site. Also, due to the limited size of a person's profile, it is essential that Twitter users double check the identity of the person whom they are communicating with, as it is easy to create a fake profile and communicate with an unknown charlatan on the service.

Social Networking Sites

Although different types of social media are often categorized as social networking sites, for the purposes of this paper, social networking sites are defined as Web-browser and smartphone accessible services that allow users to create social connections in a public or semi-public form (through the use of profiles) in order to share information updates with other site users. Wikipedia, the online user-generated encyclopedia, further expands on this definition with a number of concepts, as can be seen below [23]:

> *A social networking service is an online service, platform, or site that focuses on facilitating the building of social networks or social relations among people who, for example, share interests, activities, backgrounds, or real-life connections. A social network service consists of a representation of each user (often a profile), his/her social links, and a variety of additional services. Most social network services are Web-based and provide means for users to interact over the Internet [through] instant messaging. Online community services are sometimes considered as a social network service, though in a broader sense, social network service usually means an individual-centered service whereas online community services are group-centered. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks.*

The majority of the peer-reviewed literature on social networking is centered on issues of maintaining professionalism, ethical practices, identity, and privacy. However, given that these subjects apply to all types of social media, they are addressed in the discussion section. Cyberbullying is also a common topic of discussion of the literature in this space; however, it is out of scope for this paper.

An iconic paper by Farmer and colleagues (2009) [51] evaluated the relationship between Facebook groups and common medical conditions. They found that the most common type of groups on Facebook were centered on specific medical conditions (eg, malignant tumors), peer-to-peer support, and fundraising for support groups, organizations, and individuals. Farmer and colleagues also found that researchers used Facebook to aggregate themselves into a "network" for dissemination of their research to other researchers and health care providers. They also identified the existence of self-aggregated negative-behavior support groups, mainly centered on the promotion of excessive alcohol consumption.

Similarly, Bender and colleagues (2011) [52] found that the majority of those who use social networking sites use them to form self-aggregated interest groups. Within a single disease, breast cancer, a search on Facebook revealed over 600 support groups organized around four central themes: fundraising, awareness, marketing, and general support. General support groups were not used as an adjunct to supportive care nor did they serve as a general form of patient-to-patient support; rather, they were most often created by a user (or family member) with cancer as a means to keep friends and family members updated on their treatment and, at the same time, receive supportive feedback. Bender et al [52] also noted that their results may be skewed because they were able to analyze public groups only, which had very few user contributions as a whole. Furthermore, the technical architecture of Facebook, which makes it difficult to have a fictitious profile when compared to other (more open) social media such as Twitter, may also be responsible for skewing the data. This is a general limitation of research on social media sites—all closed profiles and private conversations cannot be evaluated unless the actual patient discloses the content of these interactions, thus this literature review did not find any formal research comparing “closed” groups on Facebook.

Another interesting use of Facebook groups occurred in Taiwan [53], where a well-known emergency physician blogger created a public group to ask his colleagues as to how they could improve patient wait times in the emergency room. In less than a month, the group grew virally, with the majority of emergency department staff from around Taiwan proposing solutions. Eventually, the group received so much attention that the Minister of Health himself (and his staff) joined the group and commented directly, using the comments from its 1500 plus group members to make policy decisions. This culminated with the minister making visits to emergency departments in ten different cities and promising to initiate a dialogue to improve funding and reduce wait times in emergency departments in collaboration with the Taiwanese Bureau of National Health Insurance.

An important facet of most social networking sites is that third-party applications (apps) can be created within these services. Third-party applications work through the integration of application programming interfaces (APIs) that allow outside software and data to be visualized. In Facebook, the most prominent of these is Zynga’s Farmville Game, which allows people to create a virtual farm and, by interacting with other Facebook friends, acquire a virtual currency that can be used to buy virtual goods, such as tractors or animals.

Similar examples within health can be found in an article by Fernandez-Luque and colleagues [54] that searched for and evaluated these “apps” within Facebook. This research found that less than 30% of listed applications were real and the remaining 70% were non-functioning “spam”. In their evaluation of the 56 working applications, Fernandez-Luque and colleagues found that these software were thematically centered on fitness and weight loss, specific health conditions (eg, diabetes) education, smoking cessation, and fundraising for health and research-related activities. The most notable included “Get up and move”, which allows people to challenge their friends to engage in physical activity and report on it after they have completed it; the American Heart Association’s “START” application, which was part of a heart portal and allowed users to answer questionnaires on cardiovascular health and upload the data to a health portal; and HealthSeeker, a diabetes management education app (see Figure 2) allowing users to learn how to better manage their diabetes and gain “points” that could be used for incentive draws in the process. Although not specifically named, two other applications were also described that allowed users to make appointments for blood donations. Only one application was made for physicians, which was used as a forum to answer patient questions.

Within the research realm, Bull and colleagues published a reflective case study that discusses the ethical questions that emerged during a Facebook-based randomized controlled trial of preventative HIV education for high-risk teens in the United States [55]. They found that maintaining ethical principles was the most difficult part of using Facebook for research. In particular, maintaining beneficence, improving knowledge and information comprehension, ensuring equity of special populations, and safeguarding confidentiality and security were the largest challenges to the study’s implementation. To overcome these problems, Bull and colleagues referred study participants from a Facebook fan page to an external website, which was congruent with the US Health Insurance Portability and Accountability Act (HIPAA) and their Institutional Review Board’s requirements. Bull and colleagues concluded by recommending that researchers who plan to collect data from social networking sites consider whether the social networking service is the appropriate vehicle for participant recruitment, that they offer multiple venues for participants to provide informed consent, and that all data are safeguarded behind secure firewalls, preferably outside the original social networking site.

Figure 2. The HealthSeeker Diabetes Education App on Facebook.

I WILL Move It! Move It!

Okay, so you bought into the reasons for being physically active. That's the easy part. The harder part is...

Size Matters!

It seems like food portion sizes have gotten a little out of control! Thanks to the "supersizing" phenomenon...

Carbs Count!

Let's clear up any confusion right now: carbs aren't bad! What

E Promote healthy eating and good nutrition

D Assist in improving or maintaining diabetes (glycemic) control

W Promote achievement and maintenance of a healthy weight

H Promote heart health and reduce the risk of heart disease

"Always leave the table a little bit hungry"

Elliott P. Joslin, M.D., Founder, Joslin Diabetes Center

Professional and Thematic Networking Sites

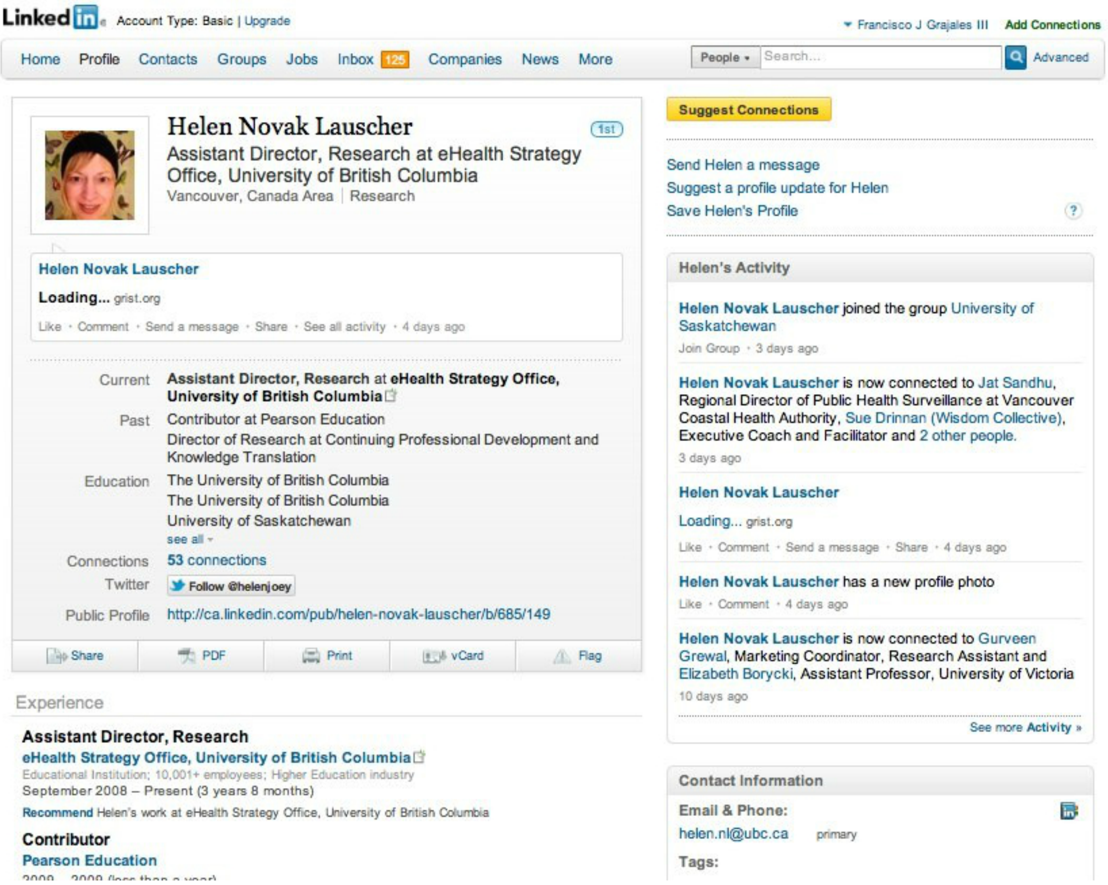
Professional networking sites are aimed solely for interactions related to a person’s professional career or business. LinkedIn is the most popular of such sites and does not solely focus on medicine or health care; it allows people to publicly display a curriculum vita along with personal and institutional affiliations. Unlike Facebook, which allows people to “friend” each other, LinkedIn uses connections, which publicly show people that have worked together or know each other. Should a person be new to the site, connections can also be used to visualize the number of degrees of separation between 2 or more people. Figure 3 displays a public profile on LinkedIn.

A number of health care specific professional networking sites also exist, including Sermo, Asklepios, Doctors’ Hangout, Ozmosis, Doc2Doc, and others, which try to recreate the intimacy of the “physicians’ lounge” in an online environment. These sites most often require the clinician to submit their credentials to a site gatekeeper, thus creating the perception of an elitist forum that is “safe” from patient’s eyes. Discussions in these sites typically range from dating in a medical environment, ethics, clinical trial and medication reviews, biostatistics, and specific treatment options. A combination of business models are also used by these sites, which vary from financial sponsorship by a professional association (eg, Asklepios by the Canadian Medical Association), advertisement, anonymized data vending to external stakeholders (eg, insurance companies, pharmaceutical companies, etc), commission on prizes offered by companies trying to solve a particular problem (eg, InnoCentive), and research by external stakeholders (eg, surveys on physician medication prescription habits).

Thematic networking sites are analogous to professional networking sites but centered on a particular theme. These include telemedicine (eg, Telehelp), informatics (eg, Health Informatics Forum), nursing (eg, SocialRN), genomics (eg, 23andMe), and patients (eg, PatientsLikeMe), among others. Of particular interest are patient thematic networking sites, as a number of these sites collect, aggregate, and visualize patient data to promote patient-driven research (research that was initiated by a patient and used to collaborate with other patients with the same or a related disease) [56].

One site that promotes patient-driven research is CureTogether. It collects a number of health metrics including weight, caloric intake, sleep, exercise, and other disease-specific indicators [57]. Although the site is not meant to constitute medical advice, it allows patients to summarize statistics on treatment efficacy, side effects, and causes of disease, ultimately helping people make more informed treatment decisions. For example, on its chronic fatigue syndrome page, CureTogether has amalgamated responses from over 1300 patients, encompassing nearly 7000 data points on effective treatments. Indeed, it is this “crowdsourced” or collective wisdom that is believed to combat single stakeholder bias. This approach remains strenuously contested by physicians and the public alike, as it is difficult to prove, in terms of accuracy and validity, that third parties have not intervened in how results are displayed to users. Figure 4 displays a summary page on different treatment effectiveness for chronic fatigue syndrome.

Figure 3. A public profile on LinkedIn.



Account Type: Basic | Upgrade

Francisco J Grajales III Add Connections

Helen Novak Lauscher 1st

Assistant Director, Research at eHealth Strategy Office, University of British Columbia

Vancouver, Canada Area | Research

Helen Novak Lauscher

Loading... grist.org

Like · Comment · Send a message · Share · See all activity · 4 days ago

Current **Assistant Director, Research** at **eHealth Strategy Office, University of British Columbia**

Past Contributor at Pearson Education

Director of Research at Continuing Professional Development and Knowledge Translation

Education The University of British Columbia

The University of British Columbia

University of Saskatchewan

see all

Connections 53 connections

Twitter Follow @helenjoey

Public Profile http://ca.linkedin.com/pub/helen-novak-lauscher/b/685/149

Experience

Assistant Director, Research

eHealth Strategy Office, University of British Columbia

Educational Institution; 10,001+ employees; Higher Education industry

September 2008 – Present (3 years 8 months)

Recommend Helen's work at eHealth Strategy Office, University of British Columbia

Contributor

Pearson Education

2009 – 2009 (less than a year)

Send Helen a message

Suggest a profile update for Helen

Save Helen's Profile

Helen's Activity

Helen Novak Lauscher joined the group University of Saskatchewan

Join Group · 3 days ago

Helen Novak Lauscher is now connected to Jat Sandhu, Regional Director of Public Health Surveillance at Vancouver Coastal Health Authority, Sue Drinnan (Wisdom Collective), Executive Coach and Facilitator and 2 other people.

3 days ago

Helen Novak Lauscher

Loading... grist.org

Like · Comment · Send a message · Share · 4 days ago

Helen Novak Lauscher has a new profile photo

Like · Comment · 4 days ago

Helen Novak Lauscher is now connected to Gurveen Grewal, Marketing Coordinator, Research Assistant and Elizabeth Borycki, Assistant Professor, University of Victoria

10 days ago

See more **Activity** »

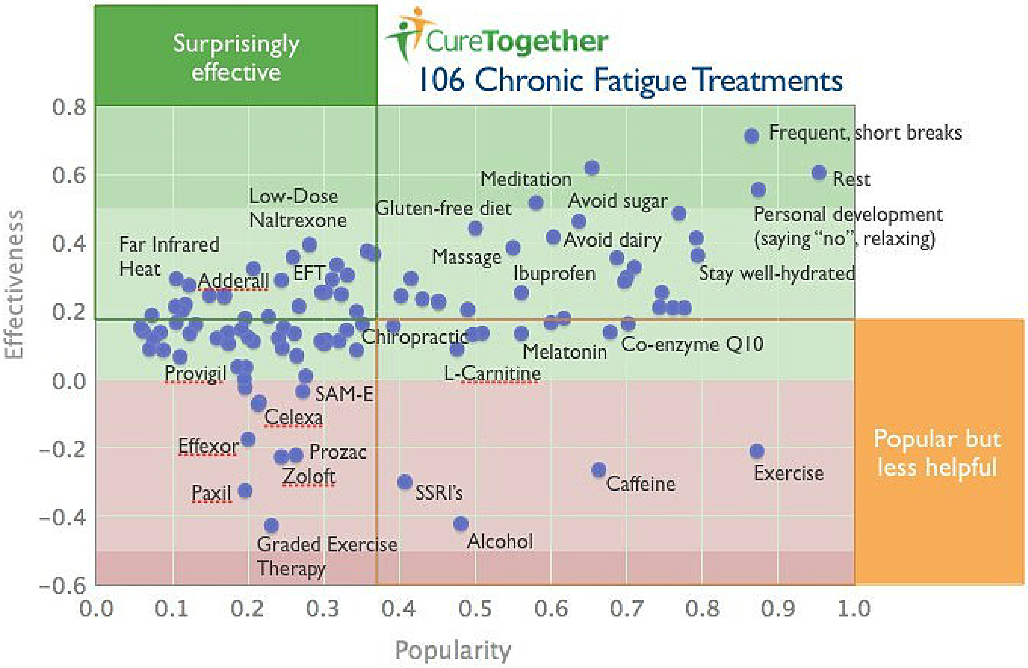
Contact Information

Email & Phone:

helen.nl@ubc.ca primary

Tags:

Figure 4. CureTogether's page on chronic fatigue syndrome.



Wikis

Wikis are easy-to-publish websites that can be quickly and easily edited by multiple users; they feature both text and multimedia content. (The term wiki was borrowed from the Hawaiian language and it stands for quick; it is a metaphor of the speed with which information can be accessed, added, and edited on a wiki.) Within medicine, the most commonly cited and used wiki is Wikipedia, which receives over 150 million page views per month, with the top 200 medical articles receiving over 100,000 monthly page views [58]. Additionally, the Health Library Wiki of Canada (a University of British

Columbia Library initiative) lists over 61 medical wikis and wikibooks available to health care professionals and patients alike.

Although Wikipedia's accuracy and completeness are often debated, as is the content of many wikis, research by Clauson and colleagues [59] compared Wikipedia's drug information with the Medscape Drug Reference. This research found that Wikipedia had very few factual errors and that it included approximately 76% of the content found in Medscape (a validated and trusted information source) [59]. Wikipedia was also found to foster quality/accuracy improvements over a period of 90 days, due to crowdsourcing, which was not the case with Medpedia, due to their more restrictive editorial process [59]. In their concluding remarks, Clauson et al noted that Wikipedia was a good starting point for health information consumers, despite being narrower in scope, less complete, and with some errors, mainly of omission, than Medscape.

In terms of its editorial membership, a recent article published by some of Wikipedia's medical editorial team (a self- and peer-selected group of WikiProject contributors who work together to improve Wikipedia) [58] noted some of the dilemmas faced on this digital space. Among the weaknesses noted is that some people and organizations (eg, the pharmaceutical industry) have used the site to advance their personal and corporate mandates. Also, Wikipedia's user architecture uses a self-selected pseudonym for authors and editors of content, which makes it very difficult to verify the expertise or credentials of the contributors. Among its strengths, Wikipedia has an effective rating system analogous to peer review. Identification of promoting an article to "Good Article Status" requires the independent review by at least 1 editor, and to "Featured Article Status", review by a group of editors. Becoming an administrator on Wikipedia is also no easy task, as user rank promotion is subject to a public voting system where the date and number of articles edited and authored are evaluated by other Wikipedians. In addition, IP addresses of the votes are cross-checked to prevent a single person from self-nomination. Wikipedia has a style manual [58], which, for example, recommends that drug dosages be excluded from the site in order to prevent patient harm. There is also an elaborate process of content verifiability if disputes arise between editors to ensure successful conflict resolution and neutrality of the information.

Other notable wiki use for public health include the World Health Organization's (WHO) International Classification of Diseases 11 update experiment [60], which opened the International Classification of Diseases system used around the world, allowing clinicians to consider and add new codes before the WHO ratified the new classification strata through its internal processes. Similarly, the Medskills wiki, which is a wikibook that compiles physical assessment techniques, allows students to learn them without having to buy expensive textbooks. Wikisurgery is a free surgical encyclopedia, and OpenWetWare features a number of laboratory procedures to facilitate the learning steps of complicated laboratory techniques (eg, a DNA polymerase chain reaction).

A large variety of wiki-like software exists on the Internet. These include Google Documents (GDocs), which is a word-processing program analogous in functionality to Microsoft Word but with the added benefit of simultaneous synchronous user collaboration and automatic document publishing as a webpage [61]. Similarly, Etherpad, is an open source, low bandwidth, massive user (>30) collaborative writing tool, particularly suitable for simultaneously working on a document during (tele)conferences and meetings with a large number of attendees, due to its minimalistic interface [62].

The use of wikis in health has significant challenges. These include attracting and maintaining a critical mass of content contributors and editors to maintain accuracy and currency of content, and dealing with wiki damage, which occurs through (human or robot) spam, link rot, deletion of information such as medication side effects (despite the ability to see a history of changes, similar to MS Word's track changes feature), and the deliberate insertion of misinformation (eg, neutraceutical companies noting that their products can cure cancer and other similar claims [58]).

Mashups

Mashups are combinations of two or more Web services that use APIs to create a new service or functionality. The term was borrowed from the music industry, where separate music tracks are combined by DJs to create a new song. The first medical mashups originated in the form of analog Geographical Information Systems, similar to John Snow's map of the cholera outbreak in London in 1854 [63]. Today, a large number of medical mashups exist (see Table 2) ranging from infection disease visualization (eg, HealthMap) to PubMed search engines, which use semantic technologies to find peer-reviewed articles that closely match an author's written block of text (eg, ETBlast3).

HealthMap is an example that illustrates both the importance and need for mashups by large organizations. Financed by the Google Foundation and supported by the Canadian Institutes for Health Research, US Centers for Disease Control, and the National Library of Medicine, HealthMap uses Fisher-Robinson Bayesian filtering to aggregate information from the WHO's Information System (WHOSIS), the Program for Monitoring Emerging Diseases (ProMED-mail) databases, Geosentinel (the global surveillance program from of the International Society of Travel Medicine), the World Organization for Animal Health, the European Centre for Disease Prevention and Control, Baidu and Baidu News, and Google News [64]. HealthMap combines a very large variety of data sources and APIs (eg, Google Maps, Google Translate, etc) to create a highly powerful information resource that can be "zoomed" all the way to relatively small geographic regions (eg, suburbs). When looking for disease outbreaks, all details are dependent on the source data, which means that while some geographic regions may display a high level of information, others may not, which is a general weakness of the site. Figure 5 displays the HealthMap mashup for North American disease outbreaks.

Table 2. Some well-known health-related mashups.

Site name/Address	Brief description
Google Earth	Typically known as a world visualization website, Google Earth features time-enabled maps in order to track worldwide flu trends by using google.com symptom search queries.
Healthmap.org	HealthMap, led by a multidisciplinary team in Boston’s Children’s Hospital, uses informal data sources for real-time world-wide disease surveillance and outbreak monitoring.
Sickweather.com	Sickweather uses a patent-pending algorithm to aggregate data from Facebook and Twitter along with self-reported data in order to forecast, track, and map a number of illnesses around the world.
Whoissick.org	Whoissick aims to provide current and local sickness information to the public. Although it was one of the first disease visualization mashups, today the site has little data and is likely to be defunct in the near future. The main reason is a lack of a community, which provides data to the site. Whoissick also does not reveal which data sources it uses to visualize disease and symptom outbreaks.
etest.vbi.vt.edu/etblast3	eTBLAST is an article search engine that looks for peer reviewed articles, such as those on PubMed, which resemble any block of text. Thus, one can write a paragraph and look for articles, which will support the premises noted. This mashup is a project of the Innovation Laboratory at Virginia’s Bioinformatics Institute.

Figure 5. North American outbreaks in the HealthMap Mashup.

Collaborative Filtering

Collaborative filtering sites are websites that allow multiple users to tag or classify and crowdsource information to create a user-based, bottom-up folksonomy (a user-generated, unstandardized taxonomy). Today, a collaborative filtering feature can be found in most blogs (eg, to classify blog posts into one or more subjects or themes), microblogs (eg, through the use of hashtags), wikis (eg, to find related articles), and media sharing sites (eg, to find similar pictures or videos). Content tags are used to facilitate information filtering and when combined with a semantic (text analysis-based) algorithm, which prevents typos from being incorporated into the folksonomy as a new classification term, are powerful data curating tools. Three prominent collaborative filtering sites are Delicious, Digg, and Connotea [20].

Digg is a social news site that allows registered users to give a “thumbs up or down” on a news story. By doing so, articles are pushed up or down on the site’s landing page, allowing readers access to the “best” content as voted by the community. Users also have a comment and follow feature, allowing them to access other user’s views and subject interests. Moreover, Digg also integrates a number of APIs, such as Facebook Connect, allowing users to share articles and their views on them to other social media sites. In health care, Digg can be particularly useful to policy makers and hospital administrators who wish to get information on the latest outbreaks and health-related news in their local community, as the site allows access to the "highest regarded sources" as voted by the Digg user community.

Delicious is a social bookmarking service that allows users to store, share, and discover Web bookmarks. Its primary allure is a user-chosen tagging system, which allows people quickly to filter through a large number of bookmarks in a short period of time. Also, due to its Web-based nature, users can access their bookmarks from any computer with Internet access. Other features include public and private bookmarks, groups, and similar and popular link suggestions, which allow for collaboration. In medicine and health care, Delicious can be used to create high quality collaborative knowledge repositories (eg, with resources from WHO, Centers for Disease Control [CDC], Health Canada, etc) that are centered on a particular topic (eg, a treatment) and can be easily accessed and by a select (or open) group of people (eg, a hospital department).

Connotea is a free online reference management site for clinicians and scientists. It allows users to share and organize their references and receive updates as to what colleagues are reading and adding in their reference libraries [65].

It should be noted that the Web traffic rating site, Alexa, shows that despite collaborative filtering sites' usefulness, they are losing popularity and market-share in attracting new and maintaining old users due to the rising integration of a tagging feature in other social networking sites.

Media Sharing Sites

Media sharing sites comprise a large palette of social media tools that are optimized for viewing, sharing, and embedding digital media on other Web services. They share a large number of attributes with other social media—profiles, friends, comments, and private messaging/sharing of content—but their success is determined by the type of content uploaded and shared. Views are often not necessarily related to the quality of the media or its accuracy, as viral content sharing may be erroneous or have poor resolution. Most often, a site's catchy title [66], amusement level (eg, a baby panda sneezing and the mother panda being scared by the baby) [67], or relevance to current affairs is what affects its "virality" [68].

Media sharing sites, nevertheless, are great resources for knowledge translation (eg, the Ken Jong CPR video [69]), community building (eg, multiple sclerosis patient-to-patient videos [70]), marketing (eg, Viagra commercials and information pamphlets [71]), research (eg, video explaining patient rights [72]), education (eg, medical skill demonstration videos and summary sheets [73]), and branding (eg, Mayo clinic patient playing piano video [74]). They are also easy to use, have no cost (for non-premium accounts), and are accessible from both desktop and mobile devices. Table 3 illustrates the different types of media sharing sites, a common example, and their description.

A number of articles have been published on the use of media sharing sites, primarily focusing on the use of audio and video podcasts for health professional education, patient-to-patient communication, and public health campaigns.

Within the patient-to-patient communication realm, Fernandez-Luque and colleagues (2009) evaluated the comments from a random sample of YouTube videos created by patients with multiple sclerosis [75]. They found that virtual communities emerged through the "comment" feature of the site, with patients responding to each other's videos, documenting the progression of their disease, and endorsing certain medications that should be used as a last resort for the treatment of a disease (eg, Tysabri). Of concern was the direct interaction of the pharmaceutical industry with patients, requesting them to contact pharmaceutical reps to become champions and public advocates for particular medications.

Similar research by Keelan and colleagues at the University of Toronto featured a characterization of available immunization information in YouTube [76]. They found that the most commonly discussed vaccine topic was childhood vaccines (accounting for 25% of the total vaccine videos) with the most specific vaccine topic being HPV (human papilloma virus). Overall, negative videos (eg, those that contradicted the Canadian Immunization Guide) were more likely to receive a higher number of views and user ratings, and accounted for approximately 50% of total YouTube immunization videos. (YouTube does not discriminate video ranking based on content unless the video violates copyright policy; in which case, it is removed. Generally, the number of times a video is viewed is the main driver behind search result rankings.)

Media sharing sites have also become encyclopedic resources. Among the most notable are the Khan Academy, which hosts over 3000 videos and practice exercises in everything from algebra to medicine and health care, and the Doctors' Channel, which hosts online videos for a variety of health care professionals featuring content about continuing medical education, medical news, and health care–related entertainments.

Negative effects from media sharing sites have also been reported [75]. YouTube, copyright infringements are common; however, copyright owners can opt in to receive a share of advertisement revenue to keep content online. Also, few child protection initiatives have been implemented on these services. For example, if one types "proana" and "thinspiration" on YouTube, over 27,000 collective videos can be recalled. These can lead individuals to cause themselves harm by applying information on how to support anorexia and bulimia, as well as finding other equally ill individuals who become supporters in maintaining a disease-prone lifestyle [77].

Finally, it should be noted that not all content available on media sharing sites is available to anyone with an Internet connection—some countries block access to these sites (eg, China). Notwithstanding, it is also important to consider that the high prevalence of mobile phones has broken the capture and upload barrier for these sites, which means that if an organization or individual is not constantly monitoring their online presence, it is easy for an individual to take a video of themselves complaining about the care they have received and upload it onto the Web, damaging an individual or a hospitals' online reputation very quickly and with little effort.

Table 3. Types and descriptions of different media sharing sites.

Media sharing site category	Example	Description
Video sharing	YouTube	Video sharing site where users can upload, view, share, and comment an unlimited number of videos in both analog and high definition resolutions.
Photo sharing	Flickr	Image and video hosting site with an online community centered on its users and the theme of uploaded photos.
Presentation sharing	SlideShare	Slide sharing site where users can upload presentations in MS PowerPoint, Keynote, Open Office, and .pdf formats.
Document sharing	Scribd	Document sharing site where users can upload different types of document, presentation, and spreadsheet formats.
Music sharing	MySpace	MySpace was the largest social networking site until 2008; however, today MySpace is primarily used as a niche media-sharing site for musicians and emerging artists, which allows them to upload and sell single music tracks and entire albums in MP3 format.
Education sharing	iTunesU	A podcasting service provided through the Apple Corp. iTunes Store which grants free and paid access to educational documents, audio, and video. Content is multidisciplinary and available from kindergarten all the way through university; it includes course lessons, lectures, labs, and lab demonstrations.
Video and images in medicine	Medting	A Web and mobile platform that allows physicians to share medical images and build clinical cases to foster inter and intra institutional collaboration.
Theme specific	The Doctors Channel	Medical video site that offers free CME, medical news, and physician education videos from experts in over 50 specialties.

Multi-User Virtual Environments and Other Social Media

Although a large number of social media sites and functionalities are likely to continue emerging, the only remaining category of social media that has not yet been discussed is Massively Multiplayer Online Games (MMOG), more recently branded as Multi-User Virtual Environments (MUVE). These 3-dimensional ecosystems are analogous to a mashup of video games and wikis, which allow users to interact with each other through a virtual representation of themselves known as an avatar. Figure 6 displays an avatar inside a virtual operating room in Second Life.

MUVEs can be classified in two general categories: general purpose and health care specific. The most prevalent general purpose MUVE is Linen Lab's Second Life, which can be used for gaming or health care education with equal ease. Health care specific MUVEs are less common, typically focusing on particular activities such as medical education (eg, CliniSpace), simulation (eg, OpenSim), and psychiatric treatment (eg, InWorld Solutions). A large body of research exists on the use of MUVEs and is summarized below, particularly focusing on SecondLife.

Historically, MUVEs evolved from early role-playing games. These were text-based and played by multiple users through networked computers; however, computer graphics today allow live rendering that "feels" quite life-like. MUVEs are programmed to simulate many aspects of "real life" in 3 dimensions; thus, when 2 avatars walk closer together, the computer's user will experience the opposite character's voice getting increasingly louder, as in real life. This effect is also mimicked graphically; that is, other avatars (and their surroundings) are rendered with increasing sharpness and become more life-like in their interactions as they get closer together.

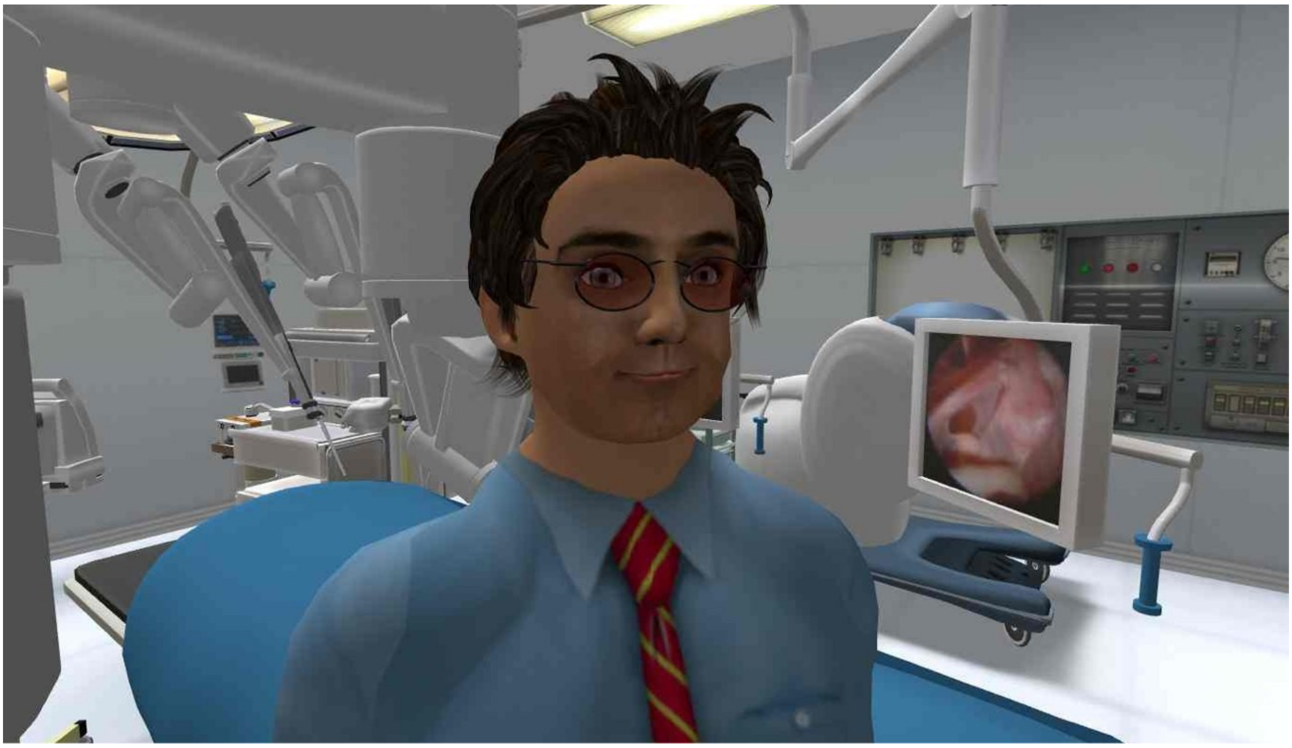
Some special-purpose MUVEs can even integrate the use of external sensors (eg, built in webcams in laptop computers) to replicate the user's emotions on their avatars (eg, smiling) [78] and experiments are already underway to incorporate scent, temperature, robotics, and even remote-controlled haptic devices. These extra gadgets have the goal of expanding current MUVE capabilities to add a "fourth dimension" (4D) [79].

Evidence for the use of MUVEs in medicine is growing rapidly with applications in health care [80] and patient education [81], epidemiology [82], mass prophylaxis simulation [83], psychotherapy [84], and research [85].

A paper by Hansen [86] has summarized the major strengths and weaknesses of these environments, which are applicable to both general purpose and health care specific MUVEs. Their strengths lie in their ability to be accessed from the comfort of a user's own home at any time and their pedagogical flexibility allowing users and content creators with knowledge of the Linden Scripting Language to design and construct a unique environment that mimics "real world" architecture. Their dynamic nature also supports collaboration at a distance, analogously to telemedicine. This is not without cost, however, as the technical barrier to entry in terms of usability often prevents and frustrates most users of these tools (eg, users have a hard time manipulating the avatar on the system and teleporting it to a virtual hospital). Other weaknesses of MUVEs also include the large amount of time required to build a 3D rendering of a physical place and the low efficiency associated with sharing text, images, and videos to an avatar, when compared with standard Web browser-based interfaces of other social media. Finally, the fact that MUVEs are often perceived as computer games, rather than serious clinical and social environments, can also affect their adoption within health care institutions.

Beard and colleagues also conducted research that surveyed health-related activities on Second Life [87]; they found 68 health-related locations. Other notable findings of this paper include the fact that research has demonstrated that using MUVEs can have real-life behavior implications. Indeed, this is the premise behind the US CDC education center on Second Life, which aims to engage visitors to influence real-life health decision-making.

Figure 6. An avatar inside a virtual operating room in Second Life.



Discussion

Ethics, Professionalism, Privacy, and Confidentiality

The potential violation of ethical standards, patient privacy, confidentiality, and professional codes of practice, along with the misrepresentation of information, are the most common contributors to individual and institutional fear against the use of social media in medicine and health care.

Equally important but less well understood is the notion of how these issues vary according to geographical and cultural norms, and how clinicians may protect themselves during Internet-based interactions. A simple example of this is the Tweetspreekuur Dutch primary care consultation service on Twitter, which is considered to be an unethical use of technology by most professional bodies who discourage or prohibit the use of social media for patient-clinician interactions altogether [13,88].

More specifically, there are varied and evolving philosophical views by professional bodies both supporting and condoning the use of social media. These contradictions are further perplexed by regional (eg, health authority) and institutional (eg, hospital) variations in policy. Fundamentally, however, the fear of the unknown appears to be a major barrier against the adoption of social media in clinical settings. This “unknown” is likely due to the conservative nature of health care institutions and practitioners, a lack of understanding of the true risks and liabilities that could result, as well as the question of whose recommendations and best practices should be followed (eg, the Canadian Medical Association [89] supports the conservative use of social media while the British Medical Association [88] and American Medical Association [13] condemn it).

Despite this uncertainty, Hrynaszkiewicz and colleagues [90] recommend that if information is posted publicly, it should not include patient identifiers (eg, patient names, insurance numbers, photos) without written consent. However, if permission is not obtained, clinicians can remain on solid ethical grounds by disclosing up to a maximum of three indirect patient identifiers (eg, sex, disease, treatment).

In general, there is a trend in the literature that recommends all clinicians to ignore patient requests sent through social media [88]. This is important because one could argue that, by not responding to these requests, clinicians are committing an Act of Omission, as there would be implied consent to respond through the medium given that the patient started the dialogue on social media (eg, Twitter).

In terms of social media policies, Boudreaux’s social media governance site [91] is the most comprehensive public database of institutional policies on the subject. As of January 2013, this resource included 219 social media policies, of which 21 were from health care institutions, including stakeholders like the Mayo Clinic, Kaiser Permanente, and Roche. In 2010, research by Kind and colleagues [92] evaluated all US accredited medical schools (n=132) for their social media presence. They found that although 95% (n=126) of American medical schools had a Facebook presence, only 13 had social media policies and only 7 of those encouraged the thoughtful and responsible use of social media.

Research has shown [93,94] that the numbers of privacy and confidentiality violations committed by physicians who use social media are small. For example, Thompson and colleagues at the University of Florida [94] evaluated 1023 student and medical residents' Facebook profiles in 2007 and 2009. They found that medical students were more likely than residents to violate privacy; however, the only privacy violations that were found were photos of medical mission trips where clinicians were interacting with patients. Even then, out of the more than 1000 profiles evaluated, only 12 ethical violations were found, which accounted for less than 2% of physician profiles. Similar research using content analysis was also conducted with self-identified physicians on Twitter by Chretien [93], who found that out of 314 physicians, each with more than 500 followers, only 3% of their total tweets could be considered unprofessional and 0.7% of them represented potential privacy violations. Nevertheless, even if violations seldom occur, health care professionals should always protect patient privacy and confidentiality as it is the ethos of the medical profession.

Information Quality

The notion of health-related information quality has been a heated topic of discussion since the mid-1990s when Internet became accessible to the public. Of concern are not only child protection and antipornography initiatives but also quackery and e-pharmacies, which often use social media for direct to consumer (DTC) advertising [95,96]. Equally important are questions of identity theft, misrepresentation of identity (eg, someone falsely claiming to be a medical doctor), and the validity of information that is provided within and through social media.

To establish the validity of the information provided, clinicians who use social media use one or more of the following tactics. First, they take pictures of themselves in a clinical environment and upload them publicly using a service such as Flickr or Picassa, so that any user with access to the Internet can see them. Second, they complete a (public) professional networking site profile, such as LinkedIn, which denotes the location and year of their medical training, professional connections and affiliations, and other credentials or interests. (It is important to remember that professional connections or "friendships" on professional networking sites may give a sense of validation from third parties under false pretenses.) Third, they apply to and are congruent with one or more information quality policy consortiums such as the Healthcare Blogger Code of Ethics (MedBloggerCode) [97] or the Health On the Net (HON) [98] Foundation's information quality initiative, which allow people to display a digital "ribbon" on their websites with a link to a third-party site that verifies compliance with their principles of information quality.

The principles of information quality, as agreed by HON and MedBloggerCode have been in existence for over a decade; however, they are still questionable because inaccurate and false information is difficult to monitor and police. Even so, most information quality "verification" bodies are reactive rather than proactive when their principles are violated. Verified websites displaying approval ribbons must typically provide the following information: (1) (professional) perspective (eg, is the blogger a cardiologist or a cardiac surgeon?), (2) confidentiality (eg, is patient privacy being protected?), (3) conflicts of interest (eg, is the writer being paid by the pharmaceutical industry?), (4) reliability (eg, are there citations to peer-reviewed material?), (5) courtesy (eg, is third-party content attributed?), (6) purpose (eg, is the purpose of the site clearly stated?), (7) justification of claims (eg, what is the level of evidence behind the information provided?), and (8) contact information (eg, are the contact details of the article author and website publisher accurate?) [97,98].

In contrast, clinicians who use social media anonymously typically use the quality of their content and the minute details provided in their rants to prove the validity of their claims and (to a point) credentials.

At the time of this study, the WHO headquarters in Geneva, Switzerland, is leading a request to the Internet Corporation for Assigned Names and Numbers (ICANN), which manages all domain names on the Internet, for a new specific .health domain. The acquisition of this domain would be strictly legislated and monitored according to quality criteria, such as those noted above, and later prioritized by a consortium of industry partners (eg, Google) to come up as the first search results when people look for health-related information [99]. Theoretically, this would improve consumer confidence with regard to the quality information from the start and improve information trust as a whole because one could validate content from social media sites directly from their Web address. However, whether this initiative will happen is a political issue that requires the support of at least 99 of 198 member states at the World Health Assembly (WHO's governing body) and will likely not be resolved in the near future.

Validated information sites have also existed since the dawn of the Internet. In social media, wikis like Medpedia (a Harvard, Stanford, University of Michigan, and UC Berkley initiative), which verify authors' credentials before allowing them to generate content, have tried to improve information quality and "validity". However, when the majority of articles by these "author verified" sources are compared with open initiatives like Wikipedia, they tend to be shorter, of equal or less quality, and have fewer references due to the restriction of users that can add and democratize the amount of content available on the Internet [100,101].

Videos and multimedia in social media create new challenges but also offer new solutions, such as steering consumers to higher quality information through peer ratings and other forms of "apomediation" [102,103].

Unanswered Questions

Developing an appropriate "standard of care" involving digital interactions, particularly those over social Web tools are likely to remain a misty ether of agreements due to the range of philosophical, cultural, social, and political values that can be found in the health sector. Professional standards, outlining whether to separate or merge clinical and personal identities are a recurring issue, as health providers have different levels of digital literacy and educational credentials (eg, should nurses have different digital interaction standards from physicians?) [104]. For example, while some professionals may deem it acceptable to use one Facebook profile for both work and personal purposes by using due diligence and monitoring their privacy settings carefully (eg, with whom they share specific information; patients vs close friends), others may lack the technical knowledge necessary to separate their personal and professional life and may have more than one profile or will avoid interactions with current and former patients altogether [95].

There is also a question of whether legal frameworks from telemedicine can or should be adapted to social media. In telemedicine, for example, the common practice is for clinicians to be licensed in the location where the patient is receiving treatment [105]. However, the Internet transcends geographical boundaries, making it nearly impossible to follow the same precedent. Even so, if the patient discloses their address of residence but is physically in a different geographical location when receiving treatment, it is debatable whether this principle should be maintained.

Other issues that need to be addressed, in terms of liability and malpractice, is whether a health care provider, layman, or digital platform should be held liable for health-related recommendations provided through social media [95]. The articulation of policies that are adaptive to the rate of newly evolving social technologies will also continue to be a challenge for decision makers. At the core is the question of whether professional organizations (eg, the British Columbia College of Physicians and Surgeons or its equivalent in other local jurisdictions) would prefer to monitor and enforce every digital interaction or whether they will grant the discretion necessary for their members to exercise their professional judgment and due diligence and undertake an investigation only when they receive a complaint.

Furthermore, there is a need for an urgent evaluation of policies by key actors (eg, Public Health Agency of Canada, the Canadian Medical Association, Provincial Ministries of Health, etc) that are responsible for safeguarding computer-mediated communication in health care. Should, for example, standardization and verification of medical licensing be implemented on the Web and be linked to local licencing bodies? If so, it would need to be operationalized in such a way that provincial colleges of physicians, nurses, and other care providers can link to a database or Web-ribbon to prove their licensed clinician status. Ensuring the highest possible safety and effectiveness of digital interactions is a mutual responsibility of industry, professional associations, and government; however, no hierarchy of responsibility and accountability presently exists. The gaps in policies must be harmonized through a multistakeholder meeting or clinicians will continue to operate in a conflicting policy environment, which may ultimately lead to legal action as a result of their social media use.

Governments also need to identify what business and data-usage models are appropriate in the health sector. For example, is it appropriate to sell patient information? Traditionally, it has been appropriate if multiple patient data are aggregated and anonymized. However, given that users seldom read the terms of service when signing up for a social tool and that they are not allowed to modify them, an ethical question remains about whether they are being de facto coerced to give their data away when joining a specific service. This issue is further complicated by the notion that interacting through social media is an increasing social expectation [106].

An additional issue is that few Web companies and social media service providers are fully transparent, from the moment a user signs up for a social service, about how they will use a user's data. Even if they are transparent in how they will use the data, the terms of service, which are legal binding documents, often change without the end user having any say in the matter or even being aware of the changes. It is important to consider whether or not users and industry would be willing to open a pay-for-privacy business model, which could potentially allow social media to become an ecosystem for safe and secure digital interactions in health care by allowing clinicians and patients to use services they already use (eg, Facebook) for a safe and ethical health care-related encounter. Alternatively, the question of whether governments should institute a legal requirement for user privacy in these sites will be important to consider in the future, as secure messaging platforms in the health care space are expensive and sometimes even subject to privacy and confidentiality breeches themselves [107].

To date, no longitudinal evaluations of the full economic effects of social media on the health domain have been conducted. Though this may be due to the novel nature of social media, such evaluations would help determine the appropriate incentives (eg, CME vs money), who should provide them, the return on investment, total cost of ownership, scalability, and long-term financial feasibility of using social media [108].

Other theoretical and pragmatic questions must also be addressed, including (but not limited to) the following: (1) Will Wikipedia and other medical wikis that use crowdsourcing and open structures of community-regulated validation become more powerful and sustainable than UpToDate-like resources that have traditionally used a small (paid) group of individuals to create clinical information summaries? (If so, what are the ethical and legal responsibilities of the Wikipedia-like actors towards health consumers?); (2) How biased are social media in providing medical information to users, and is it leading to near-infinite segregation of users around a specific belief (eg, quackery)?; and (3) What are the sociocultural, ecologic, and architectural considerations that must be contemplated over the next decade in the use of social technologies in health care?

Limitations

This study has a number of limitations. First, results were limited to the English language. Second, during the environmental scans and grey literature queries, snowballing was used, which is subject to friendly and frequent author bias. Third, due to the complexity of the data synthesis process, there was a 3-month lag between the data collection and the completion of this manuscript (despite efforts to monitor new applications and tools during this time, it is possible that new developments during this time period may have been missed, such as [109]). In light of the rapid development in this domain, this time lag from literature collection to manuscript compilation and eventual publication can be significant.

Professional Implications

Overview

This research has demonstrated the many ways that clinicians (as well as patients, health care organizations, and other related stakeholders) can use social media in health care and, as previously noted, that many ethical and legal issues remain unclear. Nevertheless, there are examples of social media demonstrating benefits to patients. Thus, short of having standards and boundaries set by health policy makers and licensing bodies to govern health professional behavior, the following four guidelines may be used to mitigate risk during such interactions over social media and most particularly, as it applies to clinicians.

Principle 1: Maintain Professionalism at All Times

Clinicians must remember and follow their institution's and professional association's social media guidelines in all digital interactions. If such bodies have not yet created a policy on the use of social media, clinicians must assume that all information exchanged is public and posted in a medium no different than a newspaper. If in doubt about whether the information to be posted is appropriate, it should not be posted. It is also essential to remember that just because a message is private (eg, a direct message on Twitter or Facebook) this does not mean that the information being exchanged is secure and protected. Clinicians and organizations may also use disclaimers to note that the information provided through social media does not indicate any form of endorsement or validation by third parties and that all views expressed are solely those of the author and not those of the institution that the clinician is affiliated with. Indeed, although disclaimers in general have no legal weight in court [78], they do inform the public of separate personal and institutional identities.

Principle 2: Be Authentic, Have Fun, and Do Not Be Afraid

The only way to create meaningful relationships over social media is to be genuine. Clinicians should not be afraid to be themselves, so long as they keep in mind Principle 1 and remember the public nature of social media, as well as who their audience is.

Principle 3: Ask for Help

People who use social media are very enthusiastic about new members joining their community; thus, clinicians should look for people with similar interests, both professional and personal, and ask for help. Attention to detail should also be placed on how people interact (eg, netiquette) and mimic the social media service and community's practices (so long as they are professional).

Principle 4: Focus, Grab Attention, Engage, and Take Action

One of the most useful models for the successful engagement of an online audience with social media is the Dragonfly Model [110]. By using the analogy of a dragonfly, which needs all four wings to work in concert, equally this model uses the following principles: (1) focus (eg, identify a single, concrete, and measurable goal for using social media), (2) grab attention (eg, make others look at content by saying or posting something interesting), (3) engage (eg, foster personal connections by discussing your interests with like-minded people), and (4) take action (eg, enable and empower others).

Conclusions

The role of social media in the health care sector is far-reaching, and this article has discussed what, where, how, and why different social media are used in a spectrum of health care–related settings. Questions and debates in terms of governing social media and its applications to medicine and health care are likely to remain contentious, or at least unclear, for some time to come.

Although research has shown that few physicians who use social media violate privacy and confidentiality standards, it is unclear as to whether it is appropriate to delegate discretion to the physicians and allow them to decide if social media is appropriate in specific medical contexts. Indeed, this is the case in the Netherlands with the primary care Twitter consultation service @tweetspreekuur, where Dutch telemedicine policies allow physicians to make the call of whether a particular technology is appropriate for patient care. Understanding which actor or actors are responsible and/or liable, as well as how ethics, confidentiality, privacy, and information quality should be managed will remain central issue that must be resolved in the coming years.

The four guidelines offered here provide a starting point for health care professionals who wish to use social media in a safe and ethical manner. However, much work remains to be done in understanding the pertinence of social media in public care when contrasted with their use in private systems where social media is principally used as a marketing technique to supplement concierge-medicine. Finally, more research will allow us to understand the synergies between social media and evidence-based practice, ultimately allowing for evidence-based policies and economic analyses on the return of investment of using social media.

Conflicts of Interest

One of the authors (GE) is 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), who was not made aware of the fact that GE was a coauthor. GE has not been involved in any editorial decisions related to this paper. GE is also producer of the Medicine 2.0 Conference Series, which deals with social media and mHealth.

Multimedia Appendix 1

Top medgadget blogs during 2010.

[PDF File (Adobe PDF File), 207KB - jmir_v16i2e13_app1.pdf]

Multimedia Appendix 2

A glossary of commonly used terms on Twitter.

[PDF File (Adobe PDF File), 218KB - jmir_v16i2e13_app2.pdf]

References

1. Mrazek D. Mayo Clinic Depression blog. URL: http://web.archive.org/web/20120816023110/http://www.mayoclinic.com/health/depression-blog/MY00151 [accessed 2013-10-27] [WebCite Cache ID 6KgqvoPNF]
2. 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]
3. 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]
4. Wikipedia. Obsessive-compulsive disorder. URL: http://en.wikipedia.org/wiki/Ocd [accessed 2013-10-27] [WebCite Cache ID 6Kgr6CMyZ]
5. Seeman N, Ing A, Rizo C. Assessing and responding in real time to online anti-vaccine sentiment during a flu pandemic. Healthc Q 2010 Sep;13 Spec No:8-15. [Medline: 20959725]
6. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
7. 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]
8. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
9. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
10. Kapp JM, Peters C, Oliver DP. Research recruitment using Facebook advertising: big potential, big challenges. J Cancer Educ 2013 Mar;28(1):134-137. [doi: 10.1007/s13187-012-0443-z] [Medline: 23292877]
11. Donahue M. Pharm Executive Blog. Patient Recruitment via social media: Lessons learned URL: http://blog.pharmexec.com/2012/02/13/patient-recruitment-via-social-media-lessons-learned/ [accessed 2013-10-27] [WebCite Cache ID 6KgrNb1Np]
12. Anderson PF, Blumenthal J, Bruell D, Rosenzweig M, Conte M, Song J. An online and social media training curricula to facilitate bench-to-bedside information transfer. 2009 Presented at: Positioning the Profession: the Tenth International Congress on Medical Librarianship; 2009; Brisbane, Australia p. 1-11.
13. American Medical Association. AMA Policy: Professionalism in the use of social media. URL: http://www.meded.umn.edu/policies/du/Social_Media_DU.pdf [accessed 2013-10-27] [WebCite Cache ID 6KgrYlwy2]
14. Kennedy S. Placenta Facebook photos: Nurse and mommy tribes see student expulsion differently. URL: http://ajnoffthecharts.com/2011/01/12/placenta-facebook-photos-nurse-and-mommy-tribes-see-student-expulsion-differently/ [accessed 2013-10-27] [WebCite Cache ID 6KgrdDpXz]
15. Schreiber WE, Giustini DM. Pathology in the era of Web 2.0. Am J Clin Pathol 2009 Dec;132(6):824-828 [FREE Full text] [doi: 10.1309/AJCPEC9FZSB4DEDH] [Medline: 19926572]
16. 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 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
17. IMIA Social Media Working Group. IMIA SMEWG at Medicine 2. URL: http://imiasocialmedia.wordpress.com/2010/11/28/imia-social-media-wg-at-medicine-2-0-nl/ [accessed 2013-10-27] [WebCite Cache ID 6KgrnXCiH]
18. Greenhalgh T, Russell J. Why do evaluations of ehealth programs fail? An alternative set of guiding principles. PLoS Med 2010;7(11):-. [doi: 10.1371/journal.pmed.1000360]
19. The Cochrane Collaboration. Social Media Resources. URL: http://www.cochrane.org/about-us/evidence-based-health-care/webliography/social-media [accessed 2013-10-27] [WebCite Cache ID 6KgrqvcJg]
20. HLWiki Canada. Social bookmarking sites. URL: http://hlwiki.slais.ubc.ca/index.php/Social_bookmarking_sites [accessed 2013-10-27] [WebCite Cache ID 6KgrtcD0H]
21. JMIR Publications. Medicine 2.0 Congress: 6th World Congress on Social Media, Mobile Apps, and Internet/Web 2.0 in Medicine, Health and Biomedical Research. URL: http://www.medicine20congress.com [accessed 2013-08-27] [WebCite Cache ID 6Kgs0Cqc1]

22. Giustini D, Grajales III FJ, Hooker D. Using social media to promote evidence based practice. URL: http://www.slideshare.net/giustinid/cochrane-social-media-workshop-2011 [accessed 2013-10-27] [WebCite Cache ID 6KgsBewP0]
23. Wikipedia. Social networking service. URL: http://en.wikipedia.org/wiki/Social_networking_service [accessed 2013-10-27] [WebCite Cache ID 6KgsEPORU]
24. Wikipedia. Professional networking device. URL: http://en.wikipedia.org/wiki/Professional_network_service [accessed 2013-10-27] [WebCite Cache ID 6KgsGxJAP]
25. Wikipedia. Mashup. URL: http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid) [accessed 2013-10-27] [WebCite Cache ID 6KgsJNzWP]
26. Wikipedia. Collaborative filtering. URL: http://en.wikipedia.org/wiki/Collaborative_filtering [accessed 2013-10-27] [WebCite Cache ID 6KgsL839t]
27. Kovic I, Lulic I, Brumini G. Examining the medical blogosphere: an online survey of medical bloggers. J Med Internet Res 2008;10(3):e28 [FREE Full text] [doi: 10.2196/jmir.1118] [Medline: 18812312]
28. Wikipedia. Paul F Levy. URL: http://en.wikipedia.org/wiki/Paul_F._Levy [accessed 2013-10-27] [WebCite Cache ID 6KgsQd8E1]
29. Chretien K, Goldman E, Faselis C. The reflective writing class blog: using technology to promote reflection and professional development. J Gen Intern Med 2008 Dec;23(12):2066-2070 [FREE Full text] [doi: 10.1007/s11606-008-0796-5] [Medline: 18830767]
30. Mayo Clinic. Sharing Mayo Clinic: Stories from patients, family, friends, and Mayo Clinic staff. URL: http://sharing.mayoclinic.org/ [accessed 2013-10-27] [WebCite Cache ID 6KgsVdkIq]
31. Gupta S. CNN Health. URL: http://sanjayguptamd.blogs.cnn.com/ [accessed 2013-10-27] [WebCite Cache ID 6KgsXiXfu]
32. Marina del Rey Hospital. Surgery superior for obese diabetic patients, studies say. URL: http://blog.marinahospital.com/surgery-superior-for-obese-diabetic-patients-studies-say/ [accessed 2013-10-27] [WebCite Cache ID 6KgsZJbDS]
33. Donahue M. Pharm Executive Blog. Patient Recruitment via social media: Lessons learned URL: http://blog.pharmexec.com/2012/02/13/patient-recruitment-via-social-media-lessons-learned/ [accessed 2013-10-27] [WebCite Cache ID 6KgsdVVzh]
34. Six Until Me. About Kerri Morrone Sparling. URL: http://www.sixuntilme.com/about/ [accessed 2013-10-27] [WebCite Cache ID 6KgsgXKpY]
35. deBronkart D. e-Patient Dave: A voice of patient engagement. URL: http://epatientdave.com/ [accessed 2013-10-27] [WebCite Cache ID 6KgsjFRZI]
36. Miller DK. The Last Post. URL: http://www.penmachine.com/ [accessed 2013-10-27] [WebCite Cache ID 6KgslWc8B]
37. Dimov V. Cases Blog - Medical and Health Blog. URL: http://casesblog.blogspot.ca/ [accessed 2013-10-27] [WebCite Cache ID 6Kgsqq0qW]
38. Pak Ng GL. Plane Crash into Hudson River. URL: http://www.flickr.com/photos/22608787@N00/3200086900 [accessed 2013-10-27] [WebCite Cache ID 6KgsuNaB2]
39. 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]
40. MedGadget. 2010 Medical Weblog Awards Sponsored by Epocrates and Lenovo: Meet the Winners!. URL: http://www.medgadget.com/2011/02/2010_medical_weblog_awards_sponsored_by_epocrates_and_lenovo_meet_the_winners.html [accessed 2013-10-27] [WebCite Cache ID 6Kgt095bY]
41. Kennedy RS. Weblogs, social software, and new interactivity on the web. Psychiatr Serv 2004 Mar;55(3):247-249. [Medline: 15001723]
42. Price B. Disseminating best practice through a Web log. Nurs Stand 2010;24(29):35-40. [doi: 10.7748/ns2010.03.24.29.35.c7627] [Medline: 20426369]
43. Poonawalla T, Wagner RF. Assessment of a blog as a medium for dermatology education. Dermatol Online J 2006;12(1):5 [FREE Full text] [Medline: 16638373]
44. Mollett A, Moran D, Dunleavy P. Using Twitter in university research, teaching and impact activities: A guide for academics and researchers. URL: http://eprints.lse.ac.uk/38489/1/Using_Twitter_in_university_research%2C_teaching_and_impact_activities_%28LSE_RO%29.pdf [accessed 2013-10-27] [WebCite Cache ID 6KgtMPLet]
45. Baumann P. 140 healthcare uses for Twitter. URL: http://philbaumann.com/140-health-care-uses-for-twitter/ [accessed 2013-10-27] [WebCite Cache ID 6KgtPLA2T]
46. George DR, Dellasega C. Use of social media in graduate-level medical humanities education: two pilot studies from Penn State College of Medicine. Med Teach 2011;33(8):e429-e434. [doi: 10.3109/0142159X.2011.586749] [Medline: 21774639]
47. Douglas F, Silvey N. Twitter journal club. URL: http://www.twitjc.com/ [accessed 2013-10-27] [WebCite Cache ID 6KgtVCxMQ]
48. McKendrick DR, Cumming GP, Lee AJ. Increased use of Twitter at a medical conference: a report and a review of the educational opportunities. J Med Internet Res 2012;14(6):e176 [FREE Full text] [doi: 10.2196/jmir.2144] [Medline: 23232765]

49. Brandenburg BJ, Jansen E. Medicine 2.0 Congress. Maastricht, the Netherlands; 2010. The proof of the Pudding: First Results of a Primary Care Consultation Service on Twitter URL: http://www.medicine20congress.com/ocs/index.php/med/med2010/paper/view/415 [accessed 2013-10-27] [WebCite Cache ID 6Kgtbrzmw]
50. Brandenburg BJ, Jansen E. Medicine 2.0 Congress. Stanford, USA; 2011. What lies around the bend? Exploring next steps in social mediaprimary care URL: http://www.medicine20congress.com/ocs/index.php/med/med2011/paper/view/575 [accessed 2013-10-27] [WebCite Cache ID 6KgteRu8y]
51. Farmer AD, Bruckner Holt CE, Cook MJ, Hearing SD. Social networking sites: a novel portal for communication. Postgrad Med J 2009 Sep;85(1007):455-459. [doi: 10.1136/pgmj.2008.074674] [Medline: 19734511]
52. Bender JL, Jimenez-Marroquin MC, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. J Med Internet Res 2011;13(1):e16 [FREE Full text] [doi: 10.2196/jmir.1560] [Medline: 21371990]
53. Abdul SS, Lin CW, Scholl J, Fernandez-Luque L, Jian WS, Hsu MH, et al. Facebook use leads to health-care reform in Taiwan. Lancet 2011 Jun 18;377(9783):2083-2084. [doi: 10.1016/S0140-6736(11)60919-7] [Medline: 21684378]
54. Fernandez-Luque L, Karlsen R, Krogstad T, Burkow TM, Vognild LK. Personalized health applications in the Web 2.0: the emergence of a new approach. Conf Proc IEEE Eng Med Biol Soc 2010;2010:1053-1056. [doi: 10.1109/IEMBS.2010.5628087] [Medline: 21097211]
55. Bull SS, Breslin LT, Wright EE, Black SR, Levine D, Santelli JS. Case study: An ethics case study of HIV prevention research on Facebook: the Just/Us study. J Pediatr Psychol 2011;36(10):1082-1092 [FREE Full text] [doi: 10.1093/jpepsy/jsq126] [Medline: 21292724]
56. Frydman GJ. Journal of Participatory Medicine. 2009. Patient-driven research: rich opportunities and real risks URL: http://ojs.jopm.org/index.php/jpm/article/view/28/18 [accessed 2013-10-27] [WebCite Cache ID 6KgtufRTC]
57. Curetogether. Curetogether surprising new data: What really helps patients with Chronic Fatigue Syndrome. URL: http://curetogether.com/blog/2011/02/03/surprising-new-data-what-really-helps-patients-with-chronic-fatigue-syndrome/ [accessed 2013-10-27] [WebCite Cache ID 6KgtyAGkh]
58. 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]
59. Clauson KA, Polen HH, Boulos MN, Dzenowagis JH. Scope, completeness, and accuracy of drug information in Wikipedia. Ann Pharmacother 2008 Dec;42(12):1814-1821. [doi: 10.1345/aph.1L474] [Medline: 19017825]
60. World Health Organization. Production of ICD-11: The overall revision process. URL: http://www.who.int/classifications/icd/ICDRevision.pdf [accessed 2013-10-27] [WebCite Cache ID 6Kgu32Vz9]
61. Archambault PM, van de Belt TH, Grajales Iii FJ, Eysenbach G, Aubin K, Gold I, et al. Wikis and collaborative writing applications in health care: a scoping review protocol. JMIR Res Protoc 2012;1(1):e1 [FREE Full text] [doi: 10.2196/resprot.1993] [Medline: 23612481]
62. The Etherpad Foundation. Features. URL: http://etherpad.org/ [accessed 2013-10-27] [WebCite Cache ID 6Kgu9g4v1]
63. Wikipedia. John Snow. URL: http://en.wikipedia.org/wiki/John_Snow_(physician) [accessed 2013-10-27] [WebCite Cache ID 6KguDC8p1]
64. HealthMap. About HealthMap. URL: http://healthmap.org/about/ [accessed 2013-10-27] [WebCite Cache ID 6KguFWRNe]
65. Connotea. About Connotea. URL: http://web.archive.org/web/20130127123905/http://www.connotea.org/about [accessed 2013-10-27] [WebCite Cache ID 6KguJwmUe]
66. Blendtech. Will it blend? - Ipad. URL: http://www.youtube.com/watch?v=lAl28d6tbko [accessed 2013-10-27] [WebCite Cache ID 6KguMCXQs]
67. Costello J. Baby Panda Sneezing. URL: http://www.youtube.com/watch?v=EAcdvmnZ_GM&feature=fvst [accessed 2013-08-27] [WebCite Cache ID 6KguO4gcp]
68. Britain's got talent. 2009. URL: http://www.youtube.com/watch?v=RxPZh4AnWyk [accessed 2013-10-27] [WebCite Cache ID 6KguQXmEn]
69. YouTube. Jen Jong CPR Video. URL: http://www.youtube.com/watch?v=n5hP4DIBCEE [accessed 2013-10-27] [WebCite Cache ID 6KgvsLcdn]
70. YouTube. Tysabri Patient Video. URL: http://www.youtube.com/watch?v=1QyaebKEisg [accessed 2013-10-27] [WebCite Cache ID 6Kgvu5vFM]
71. YouTube. Viagra Commercial. URL: http://www.youtube.com/watch?v=rE0up432ohY [accessed 2013-10-27] [WebCite Cache ID 6KgvvFuGE]
72. YouTube. Patient Rights. URL: http://www.youtube.com/watch?v=-tilcdyo8No [accessed 2013-10-27] [WebCite Cache ID 6KgvwOPRC]
73. YouTube. Suture Techniques for Skin Closure. URL: http://www.youtube.com/watch?v=-ZWUgKiBxfk [accessed 2013-10-27] [WebCite Cache ID 6Kgvyp5fJ]
74. YouTube. Mayo Clinic Piano Couple. URL: http://www.youtube.com/watch?v=RI-l0tK8Ok0 [accessed 2013-10-27] [WebCite Cache ID 6Kgw0YHfd]
75. Fernandez-Luque L, Elahi N, Grajales FJ. An analysis of personal medical information disclosed in YouTube videos created by patients with multiple sclerosis. Stud Health Technol Inform 2009;150:292-296. [Medline: 19745316]

76. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: a content analysis. JAMA 2007 Dec 5;298(21):2482-2484. [doi: 10.1001/jama.298.21.2482] [Medline: 18056901]
77. Syed-Abdul S, Fernandez-Luque L, Jian WS, Li YC, Crain S, Hsu MH, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
78. Steglman I. Product Demo and Personal Communication. 2010 Mar 02. InWorldSolutions: Private and Controlled Avatar-Based Virtual Worlds for Therapy, Training and Education URL: http://www.inworldsolutions.net/ [accessed 2013-12-22] [WebCite Cache ID 6M4PN68vt]
79. White G. Second life for the visually impaired. URL: http://blindsecondlife.blogspot.ca/2008/01/haptics.html [accessed 2013-10-27] [WebCite Cache ID 6KguXRzN7]
80. Conradi E, Kavia S, Burden D, Rice A, Woodham L, Beaumont C, et al. Virtual patients in a virtual world: Training paramedic students for practice. Med Teach 2009 Aug;31(8):713-720. [Medline: 19811207]
81. Yellowlees PM, Cook JN. Education about hallucinations using an internet virtual reality system: a qualitative survey. Acad Psychiatry 2006;30(6):534-539. [doi: 10.1176/appi.ap.30.6.534] [Medline: 17139026]
82. Gordon R, Björklund NK, Smith RJ, Blyden ER. Halting HIV/AIDS with avatars and havatars: a virtual world approach to modelling epidemics. BMC Public Health 2009;9 Suppl 1:S13 [FREE Full text] [doi: 10.1186/1471-2458-9-S1-S13] [Medline: 19922683]
83. Yellowlees P, Cook JN, Marks SL, Wolfe D, Mangin E. Can virtual reality be used to conduct mass prophylaxis clinic training? A pilot program. Biosecur Bioterror 2008 Mar;6(1):36-44. [doi: 10.1089/bsp.2007.0038] [Medline: 18386971]
84. 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]
85. Dean E, Cook S, Keating M, Murphy J. Journal of Virtual Worlds Research. Does this avatar make me look fat? Obesity and interviewing in Second Life URL: http://journals.tdl.org/jvwr/index.php/jvwr/article/view/621 [accessed 2013-10-27] [WebCite Cache ID 6KgundtVl]
86. 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]
87. 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]
88. British Medical Association. Using social media: practical and ethical guidance for doctors and medical students. URL: http://www.medschools.ac.uk/SiteCollectionDocuments/social_media_guidance_may2011.pdf [accessed 2013-10-27] [WebCite Cache ID 6Kguw5Kdu]
89. Canadian Medical Association. Social media and Canadian physicians - issues and rules of engagement. URL: http://www.cma.ca/socialmedia [accessed 2013-10-27] [WebCite Cache ID 6KguwJiyV]
90. Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. Trials 2010;11:9 [FREE Full text] [doi: 10.1186/1745-6215-11-9] [Medline: 20113465]
91. Bordeaux C. Policy database. URL: http://socialmediagovernance.com/policies.php [accessed 2013-10-27] [WebCite Cache ID 6Kgv3uwEr]
92. Kind T, Genrich G, Sodhi A, Chretien KC. Social media policies at US medical schools. Med Educ Online 2010;15:- [FREE Full text] [doi: 10.3402/meo.v15i0.5324] [Medline: 20859533]
93. Chretien KC, Azar J, Kind T. Physicians on Twitter. JAMA 2011 Feb 9;305(6):566-568. [doi: 10.1001/jama.2011.68] [Medline: 21304081]
94. Thompson LA, Black E, Duff WP, Paradise Black N, Saliba H, Dawson K. Protected health information on social networking sites: ethical and legal considerations. J Med Internet Res 2011;13(1):e8 [FREE Full text] [doi: 10.2196/jmir.1590] [Medline: 21247862]
95. Dzenowagis J, Clauson K, Grajales III FJ. Ethics 2.0: Implications of connected health. URL: http://www.medicine20congress.com/ocs/index.php/med/med2009/paper/view/273 [accessed 2013-10-27] [WebCite Cache ID 6KgvJLo72]
96. Liang BA, Mackey TK. Prevalence and Global Health implications of social media in direct-to-consumer drug advertising. J Med Internet Res 2011;13(3):e64 [FREE Full text] [doi: 10.2196/jmir.1775] [Medline: 21880574]
97. Healthcare Blogger Code of Ethics. URL: http://web.archive.org/web/20110302232738/http://www.creativetriage.com/tag/healthcare-blogger-code-of-ethics/ [accessed 2013-10-27] [WebCite Cache ID 6KgvPeOjw]
98. Health on the Net Foundation. URL: http://www.hon.ch/HONcode/Pro/ [accessed 2013-10-27] [WebCite Cache ID 6KgvTGMP8]
99. Brown P. WHO calls for a health domain name to help consumers. BMJ 2002 Mar 9;324(7337):566 [FREE Full text] [Medline: 11884314]
100. Wikipedia. Diabetes. URL: http://en.wikipedia.org/wiki/Diabetes [accessed 2013-10-27] [WebCite Cache ID 6KgvXe25b]
101. Medpedia. Diabetes. URL: http://web.archive.org/web/20130121111320/http://wiki.medpedia.com/Diabetes [accessed 2013-10-27] [WebCite Cache ID 6Kgva4dH7]

102. Gabarron E, Fernandez-Luque L, Armayones M, Lau AY. Identifying Measures Used for Assessing Quality of YouTube Videos with Patient Health Information: A Review of Current Literature. Interact J Med Res 2013;2(1):e6 [FREE Full text] [doi: 10.2196/ijmr.2465] [Medline: 23612432]
103. Eysenbach G. Credibility of Health Information and Digital Media: New Perspectives and Implications for Youth. In: Metzger MJ, Flanagin AJ, editors. Digital Media, Youth, and Credibility. Cambridge, MA: The MIT Press; 2008:123-154.
104. Moubarak G, Guiot A, Benhamou Y, Benhamou A, Hariri S. Facebook activity of residents and fellows and its impact on the doctor-patient relationship. J Med Ethics 2011 Feb;37(2):101-104. [doi: 10.1136/jme.2010.036293] [Medline: 21160080]
105. Jacobson PD, Selvin E. Licensing telemedicine: the need for a national system. Telemed J E Health 2000;6(4):429-439. [doi: 10.1089/15305620050503915] [Medline: 11242552]
106. Deloitte. Social media and mobile devices are raising the bar on HR service delivery. URL: http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/us_consulting_hctrends2012_socialmobile_021512.pdf [accessed 2013-10-27] [WebCite Cache ID 6KgvgCUM7]
107. Collmann J, Cooper T. Breaching the security of the Kaiser Permanente Internet patient portal: the organizational foundations of information security. J Am Med Inform Assoc 2007;14(2):239-243 [FREE Full text] [doi: 10.1197/jamia.M2195] [Medline: 17213500]
108. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009;28(2):361-368 [FREE Full text] [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
109. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
110. Aker J, Smith A. The dragonfly model. URL: http://www.dragonflyeffect.com/blog/model/ [accessed 2013-10-27] [WebCite Cache ID 6KgvnJ17C]
111. Twitter. The Twitter glossary. URL: https://support.twitter.com/articles/166337-the-twitter-glossary [accessed 2013-10-27] [WebCite Cache ID 6KgvoMWGS]
112. HL Wiki Canada. Top Fifty (50) Medical Wikis You Might (Want to) Know. URL: http://hlwiki.slais.ubc.ca/index.php/Top_Fifty_(50)_Medical_Wikis_You_Might_(Want_to)_Know [accessed 2013-12-12] [WebCite Cache ID 6LoLvtQd1]

Abbreviations

API: application programming interface
CDC: Centres for Disease Control
CME: continuing medical education
DNA: deoxyribonucleic acid
DTC: direct to consumer
GDocs: Google documents
HIPAA: Healthcare Insurance Portability and Accountability Act
HIV: human immunodeficiency virus
HLCanada: Health Librarianship Wiki of Canada
HON: Health on the Net
HPV: human papilloma virus
ICANN: International Corporation for Assigned Names and Numbers
MedBloggerCode: Healthcare Blogger Code of Ethics
MMOG: massively multiplayer online games
MUVE: multi-user virtual environment
RSS: really simple syndication
SMS: short message service
WHO: World Health Organization
WHOSIS: World Health Organization Information System

Edited by H Potts; submitted 27.08.13; peer-reviewed by R Booth, B Boateng, A Holzinger; comments to author 13.09.13; revised version received 29.10.13; accepted 04.11.13; published 11.02.14.

Please cite as:
Grajales III FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G
Social Media: A Review and Tutorial of Applications in Medicine and Health Care
J Med Internet Res 2014;16(2):e13
URL: http://www.jmir.org/2014/2/e13/
doi:10.2196/jmir.2912
PMID:24518354

Original Paper

Understanding Messaging Preferences to Inform Development of Mobile Goal-Directed Behavioral Interventions

Frederick Muench[1,2], PhD; Katherine van Stolk-Cooke[2], BA; Jon Morgenstern[1,2], PhD; Alexis N Kuerbis[1,2], PhD; Kendra Markle[3], PhD

[1]CASPIR, Department of Psychiatry, Columbia University College of Physicians & Surgeons, New York, NY, United States
[2]CASPIR, Mental Health Services & Policy Research, Research Foundation of Mental Hygiene, New York City, NY, United States
[3]Stanford Prevention Lab, Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
Frederick Muench, PhD
CASPIR
Department of Psychiatry
Columbia University College of Physicians & Surgeons
3 Columbus Circle 1404
New York, NY, 10017
United States
Phone: 1 212 974 0547
Fax: 1 212 974 0530
Email: fm2148@columbia.edu

Abstract

Background: Mobile messaging interventions have been shown to improve outcomes across a number of mental health and health-related conditions, but there are still significant gaps in our knowledge of how to construct and deliver the most effective brief messaging interventions. Little is known about the ways in which subtle linguistic variations in message content can affect user receptivity and preferences.

Objective: The aim of this study was to determine whether any global messaging preferences existed for different types of language content, and how certain characteristics moderate those preferences, in an effort to inform the development of mobile messaging interventions.

Methods: This study examined user preferences for messages within 22 content groupings. Groupings were presented online in dyads of short messages that were identical in their subject matter, but structurally or linguistically varied. Participants were 277 individuals residing in the United States who were recruited and compensated through Amazon's Mechanical Turk (MTurk) system. Participants were instructed to select the message in each dyad that they would prefer to receive to help them achieve a personal goal of their choosing.

Results: Results indicate global preferences of more than 75% of subjects for certain types of messages, such as those that were grammatically correct, free of *textese*, benefit-oriented, polite, nonaggressive, and directive as opposed to passive, among others. For several classes of messages, few or no clear global preferences were found. There were few personality- and trait-based moderators of message preferences, but subtle manipulations of message structure, such as changing "Try to…" to "You might want to try to…" affected message choice.

Conclusions: The results indicate that individuals are sensitive to variations in the linguistic content of text messages designed to help them achieve a personal goal and, in some cases, have clear preferences for one type of message over another. Global preferences were indicated for messages that contained accurate spelling and grammar, as well as messages that emphasize the positive over the negative. Research implications and a guide for developing short messages for goal-directed behaviors are presented in this paper.

(J Med Internet Res 2014;16(2):e14) doi:10.2196/jmir.2945

KEYWORDS

mHealth; text messaging; behavioral health; preferences; linguistics; tailoring; participatory design; agile design

Introduction

Over the past decade, mental health researchers have sought to harness popular contemporary technologies, such as computers and mobile phones, in order to develop effective interventions for a range of medical and behavioral problems. The widespread availability and real-time potential of mobile phone-based short message service (SMS) has made SMS interventions an attractive and promising subject of investigation within this area. Numerous studies have shown that SMS interventions can improve outcomes across a variety of physical and mental health disorders [1-3]. SMS interventions have yielded small to moderate effects against no treatment controls, with text messaging for smoking cessation and HIV medication adherence yielding the largest effects [1]. Recent reviews suggest that the inclusion of SMS and other prompts improves the effects of Web-based interventions and highlights the value of mobile messaging to enhance many types of interventions [4]. Moreover, brief text-based interventions of 140 characters generalize to interventions that integrate Twitter and smart phone applications, for which similar character limits apply.

The content of SMS interventions has typically been based upon prevailing global behavior change theories, such as the transtheoretical model of behavior change, social cognitive theory [5], or, more recently, specific theory-based mobile intervention mapping techniques [6]. While these behavior theories represent a foundation for present and future research on mobile interventions, many other important intervention design features have been neglected. Some digital intervention development studies have focused on human-computer interaction [7] and persuasive design features [8] that emphasize subtle differences in nonspecific intervention components like tone, design, and structure to increase user engagement in the intervention.

The tone and structure of a message can have an impact on user receptivity and engagement in an intervention, as each point of contact is an opportunity to engage the end user. A few pioneering studies have examined how message framing impacts intervention outcomes or adherence to interventions. For example, Bickmore and colleagues [9] found that using empathy in computer interactions led to a more positive user experience, while perceived politeness of task interruptions by a virtual agent predicted long-term adherence. Yet another study found that individuals exercised more when their virtual agent was serious rather than playful [10], indicating that the overall tone of an intervention can have an impact on discrete behavioral outcomes and receptivity.

There is extensive literature on the benefits of tailoring computer-based intervention content, preferences, and feedback to individual users for health outcomes across conditions [11]. Intervention tailoring increases receptivity, memory for messages, self-relevance, and self-referential processing of information for specific subgroups [12,13]. Some of the earliest work on tailoring revealed that the ideal message type (eg, picture vs text) varies based on an individual’s need for cognition [14,15], indicating that individuals process content differently and will be more receptive to some forms of message presentation than others. Despite the wealth of research on intervention tailoring, there has been almost no research to suggest which types of individuals prefer which types of message structure and content. For example, do older individuals have more difficulty processing *textese* than younger individuals? Are women more receptive to emoticons in health messages than men [16]? To date, no research has examined how demographic variables may differentially impact preferences for goal-directed short text-based interventions.

Within the general intervention development field, several development studies have used focus groups and post-pilot interviewing to examine preferences for certain types of messages. For example, participants in an SMS intervention to promote weight loss disliked the inclusion of *textese* (eg, How r u feeling 2day?) and passive language in health messaging [17]. Multiple studies across a variety of health topics have further indicated that users prefer messages that are positive in tone or benefit-oriented, brief, and direct [17,18]. Moving beyond user preferences, some recent research has specifically focused on the impact of the underlying characteristics of text on readability and retention [19,20]. For example, Leroy and colleagues [20] revealed that, although low noun-phrase complexity was perceived by users as the simplest, grammatical manipulations had little impact on the readability of the content. These studies collectively highlight the importance of understanding the impact these message features can have on user receptivity.

A useful, cost-effective method for collecting this information is rapid and iterative user preference or beta testing using quantitative methods to combat the limitations of qualitative testing. These methods have been used often in consumer research to compile data on user engagement [21] or for public health campaign engagement [22]. In our previous work developing an SMS intervention for addiction continuing care, we compared preferences for benefit-oriented vs consequence-oriented messaging and found that individuals generally preferred benefit-oriented messages, but that message preference was moderated by the perceived benefits of being drug-free [23]. This research reinforced the tailoring outcome research on the importance of congruence between motivational processes and message framing. However, along with the consumer preferences research, it also revealed that preference research may be a useful tool for initial intervention development work in resource-limited environments.

This study examined preferences for a range of text messages designed to foster goal-directed behaviors. Text messages were displayed in mirrored dyads to present participants with variations in syntax and language, tone, locus of authority, and grammatical person. The aim of this study was to determine whether any global messaging preferences existed and how certain characteristics moderate those preferences in an effort to inform the development of mobile messaging interventions. In addition, we employed iterative design techniques to assess how subtle changes in messages affected preferences from one sample to the next. This study was approved by the New York State Psychiatric Institute Institutional Review Board (NYPSI IRB) and was part of the pilot intervention development work for a mobile adaptive alcohol intervention.

Methods

Recruitment

Participants were recruited online through Amazon.com, Inc.'s online labor market, Amazon Mechanical Turk (MTurk). MTurk is a communication platform through which *workers* can be contracted to perform tasks that require human intelligence (eg, consumer surveys or beta testing) in exchange for compensation by the *requesters* who published the tasks. These tasks—called *human intelligence tasks* (HITs)—can range from one brief question to a 30-minute survey. Over the last few years, MTurk has been used for social sciences research with results similar to other sampling methods when certain validity checks were included in the design [24].

Study Eligibility

MTurk worker qualifications for this study included a HIT approval rate of 95% or greater out of at least 500 completed HITs. This ensured a sample of workers whose work on previous HITs had been consistently deemed acceptable by other requesters, as well as a sample who demonstrated an appropriate degree of computer and Internet literacy. The subject pool was further limited to participants who were located in the United States. Workers who met these qualifications could view our HIT, titled *Answer a survey about your text message preferences*, and published through our requester account, *Columbia University Research*. Eligible workers could follow a Web link to an external, Web-based survey hosted by Survey Monkey, which has been used as a survey host in numerous research studies. Prior to completing the survey, participants completed a brief consent form for anonymous survey-based research, which also provided investigator and IRB contact information. In the consent form, participants were informed that the study's aim was to understand the types of text messages they would prefer to receive when trying to achieve a personal goal. Once participants completed the survey, they were provided with a survey code to enter into their MTurk account to await requester review and compensation. Only participants referred through MTurk received compensation.

For the purposes of maintaining anonymity, we could not link the survey to the participants' MTurk accounts, but included several a priori validity checks for anonymous survey research in both the survey and our MTurk requester account. These validity checks were included in accordance with the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [25]. We published the survey four times, counterbalancing the multiple choice options and reverse ordering questions. In Survey Monkey, our safeguards included blocking IP addresses once the survey was opened by a worker in order to bar them from retaking it, omitting responses of users who did not type cogent responses to open-ended questions, and/or gave conflicting answers to a duplicated message preference question. Although we were unable to match an individual worker to his or her survey responses, we were able to view the total amount of time each worker spent on the HIT in MTurk. As our survey should take a minimum of 6 minutes to complete, we rejected the work of participants who spent fewer than 6 minutes completing it.

Participants

In total, 452 participants took one of four message preference surveys. Of those, 98 were not included in this paper because they were not located in the United States. These participants were primarily located in India and will be discussed in another paper. Of the 354 US participants, 277 were included in the final sample: 58 were excluded due to conflicting responses to identical but counterbalanced items, 9 due to missing or illegible goals, and 10 due to survey completion in under 6 minutes.

Assessments

The assessment contained approximately 90 items, which were presented in groups of approximately 8 items per screen. Participants were asked to supply a personal goal they would like to achieve and to choose one of two messages in each dyad that they would prefer to receive to help them achieve that goal. These goals did not have to be health related. Participants were told that their goal could be anything from exercise to flossing more to being more assertive, and that there were no wrong goals. There were approximately 70 message dyads in 22 groupings. Each grouping typically consisted of three dyads. Message dyads were based primarily on pre-existing motivational and behavior change content and linguistic differences in message presentation derived from previous messaging studies, public health messaging campaigns, and our own experience writing messages. Based on these existing messages, we developed a corresponding, mirrored message to test a specific preference. For example, if a message included the word *you*, we then created a mirrored message with the word "we" in place of "you". Organizational constructs for message design, message groupings and descriptions, an example dyad for each grouping, and the number of dyads per grouping are presented in Table 1. In addition, we included several single item semantic differential dyads on personality or disposition. These included face valid dichotomous semantic differentials (eg, I tend to be a sad person/I tend to be a happy person; I get along well with others/I have trouble getting along with others; I often get frustrated with the behavior of others/I don't let what others do bother me much).

Table 1. Message groupings.

Organizing Principle	Dyad Grouping	Grouping Description	Dyad Example	Dyads/ Grouping
Gain Framing vs Loss Framing				
	Smiley Emoticon vs Sad Emoticon			
		Smiley Emoticon messages contain a smiley face to make the content gain-framed.	Don’t give up :-)	3
		Sad Emoticon messages contain a sad-face to make the content loss-framed.	Don’t give up :-(	
	Benefit-Oriented vs Consequence-Oriented			
		Benefit-Oriented messages consist of language that is gain-framed.	Close your eyes – imagine the benefits of changing.	3
		Consequence-Oriented messages consist of language that is loss-framed.	Close your eyes – imagine the consequences if you don’t change.	
Personal/Emotional Emphasis				
	Coaching vs Uncoached Direction			
		Coaching messages contain a direction or recommendation with positively framed emotional emphasis.	You’ve been doing great, don’t quit now.	3
		Uncoached Direction messages contain a direction or recommendation with no additional emphasis.	The most important thing you can do to reach your goal is not give up.	
Goal-Setting and Task Performance				
	Implementation Intention vs General Goal			
		Implementation Intention messages consist of an if-then plan to trigger a specific action.	If I start to get down on myself, I will think of all my previous successes.	3
		General Goal messages consist of an open-ended, nonspecific if-then plan.	If I start to get down on myself, I will do something to make me feel better.	
Locus of Control Theory				
	Intrinsic Locus of Control vs Extrinsic Locus of Control			
		Intrinsic Locus of Control messages emphasize an internal locus of control over goal attainment.	You are responsible when you don’t meet your goal.	4
		Extrinsic Locus of Control messages emphasize the degree to which external factors influence goal attainment.	Many different aspects of your environment play a role when you don’t meet your goal.	
Spelling and Grammatical Manipulations				
	Correct Grammar vs Grammatical Errors			
		Correct Grammar messages contain no grammatical errors.	If you accept where you are now, you’re way ahead of the pack.	3
		Grammatical Error messages contain grammatical errors.	If you accept where you are now you’re way ahead of the pack.	
	Textese* vs *Non-Textese			
		Textese messages utilize the spelling abbreviations common to text messaging.	u have changed b4, u can meet ur goals today. b who u r.	3
		Non-*Textese* messages	You have changed before, you can meet your goals today. Be who you are.	
Manipulations of Visible Emphasis				
	Single Punctuation vs Multiple Punctuation			

Organizing Principle	Dyad Grouping	Grouping Description	Dyad Example	Dyads/ Grouping
		Single Punctuation messages utilize only a single punctuation mark between phrases or clauses.	Reinvent yourself!	4
		Multiple Punctuation messages utilize multiple punctuation marks between phrases or clauses for emphasis.	Reinvent yourself!!!	
	Smiley Emoticon vs No Emoticon			
		Smiley Emoticon messages contain a smiley face to enhance a friendly or positive tone.	You are on the right track :-) just keep going!	3
		No Emoticon messages contain the same language as their Smiley Emoticon counterparts, but do not include an emoticon.	You are on the right track – just keep going!	
	CAPS (capitalization) Emphasis vs No Visible Emphasis			
		CAPS Emphasis messages contain at least one world that is spelled in all capital letters for emphasis.	When it comes to the negative consequences of a bad habit, you are NOT the exception.	4
		No Visible Emphasis messages do not include any all-caps words.	When it comes to the negative consequences of a bad habit, you are not the exception.	
Manipulations of Voice, Person or Origin				
	“I” Statement vs “We” Statement			
		“I” Statement messages employ a singular first person point of view.	Changing can be hard: I promise it will get better.	4
		“We” Statement messages employ a plural first person (or collectivist) point of view.	Changing can be hard: we promise it will get better.	
	“You” Statement vs “We” Statement			
		“You” Statement messages employ a singular second person point of view.	Your past should motivate you to change – not paralyze you!	
		“We” Statement messages employ a plural first person (or collectivist) point of view.	Our pasts should motivate us to change – not paralyze us!	
	Male Quote vs Female Quote			
		Male Quote messages consist of a quote from a famous man.	“When it is darkest, men see the stars.” Ralph Waldo Emerson	2
		Female Quote messages consist of a quote from a famous woman.	“I like the night. Without the dark, we’d never see the stars.” Stephanie Meyer	
	Cited vs Uncited			
		Cited messages refer to a source/sources of the information presented.	Studies show that simply visualizing your future actions makes them more likely to come true!	3
		Uncited messages provide no point of reference for the information presented.	Simply visualizing your future actions makes them more likely to come true!	
Manipulations of Tone				
	Direction vs Passive			

Organizing Principle	Dyad Grouping	Grouping Description	Dyad Example	Dyads/ Grouping
		Direction messages express a command.	Think about what you will lose if you give up on your goals.	3
		Passive messages express a suggestion in a passive or non-urgent tone.	It could be helpful to think about what you will lose if you give up on your goals.	
	Statement vs Question			
		Statement messages utilize declarative language.	Committing to your goals today will help you in the long-run.	4
		Question messages utilize interrogative language.	How will committing to your goals today help you in the long-run?	
	Aggression vs Nonaggression			
		Aggression messages utilize a confrontational or shaming tone.	Do you seriously think that blaming others will help you change for the better?	3
		Nonaggression messages utilize a non-confrontational tone.	Blaming others probably won’t help you change for the better.	
	Polite vs Non-Polite			
		Polite messages include words such as *please* and *thank you*.	Please text us to let us know if you received this message.	2
		Non-Polite messages do not include words such as *please* and *thank you*.	Text us to let us know if you received this message.	
	Directive vs Nondirective Statement			
		Directive messages contain an imperative statement within the context of a time frame.	Call a friend to help you feel better as soon as you have a free moment.	3
		Nondirective Statement messages offer suggestions with no direction or time-sensitive context.	Going out with friends is a good idea to help you feel better.	
	Humor vs Gravity			
		Humor messages include a joke or playful tone to suggest levity.	Why did the chicken cross the road? Because it knew that action creates change.	2
		Gravity messages are serious in tone and do not contain playful or jocular language.	Action creates change.	
Symbolic Language				
	Metaphor vs Literal			
		Metaphor messages contain symbolic imagery.	When you reach the end of your rope, tie a knot and hang on.	5
		Literal messages present content in plain terms.	When you feel like giving up, keep going until it passes.	
Brevity vs Added Meaning				
	Short vs Long			
		Short messages contain as little content as possible to convey meaning.	Your actions define you.	3
		Long messages are designed to convey additional meaning.	Your actions define you: the world looks at you differently when you act differently.	

The survey was published on MTurk a total of 4 times. After the data from the first 2 survey publications was downloaded and analyzed, a number of message dyads and groupings were removed if there appeared to be a clear consensus in preference among participants (eg, the Smiley Emoticon vs Sad Emoticon grouping). New dyads and categories were then added to the survey for publication on MTurk the third and fourth time. These revisions account for the differences in the sample size for many of the message groupings examined. The content of 3 message dyads within 2 groupings was slightly altered over the course of the study in order to correct for vagueness, disproportionately weighted language, or language that did not accurately reflect the general profile of a message grouping. Specifically, dyad #3 in the Directive vs Passive grouping and dyads #2 and #3 in the Statistic vs Anecdote grouping were altered. Ultimately, the Statistic vs Anecdote grouping was excluded from the main findings due to the researchers' concern that the grouping as a whole was unsound. Therefore, only differences based on the alterations made in dyads #2 and 3 in this grouping are reported. Goals were coded into three broad categories based on their subject matter: physical health and well-being, competence and mastery, and personal fulfillment. Goals within these categories were then subcoded into more specific groupings as follows. In the physical health and well-being category, goals were subcoded as weight loss, fitness, nutrition, smoking cessation, sleep health, or personal hygiene goals. In the competence and mastery category, goals were subcoded as professional, academic, financial, or personal goals. In the personal fulfillment category, goals were subcoded as emotional, social, or spiritual goals. Finally, we included process rulers related to one's self-selected goal such as goal importance, benefits of meeting that goal, and goal efficacy, which have been used in previous research [26].

Data Analysis

A dichotomous variable for preferences within each dyad grouping was created based on a participant's majority preference for messages in that dyad (ie, at least 2/3 or 3/4 messages chosen). If a category included four messages, individuals who chose two messages of each type (50/50 preference) were removed from analysis. Moderator analysis was conducted using chi-square analysis and comparative percentages are reported.

Results

Overview

Demographics are presented in Table 2. Overall, the sample was predominately white and middle aged with at least a high school degree, and 130 out of 277 participants (47.1%) were working full-time. Most had phone plans with SMS capabilities, and 197 out of 277 (71.1%) had unlimited texting plans.

Table 2. Demographics (n=277).

Variable		n (%)
Age (years)		
	18-30	113 (40.8)
	31-40	90 (32.5)
	41-older	74 (26.7)
Gender (% female)		156 (56.5)
Race		
	Black	19 (6.8)
	White	225 (81.1)
	Asian	20 (7.1)
	Other	12 (5.0)
Ethnicity		
	Hispanic	22 (8.2)
Education level		
	High School or GED	33 (12.1)
	Some College	77 (28.2)
	College Degree	123 (45.1)
	Graduate Degree	40 (14.7)
Employment status (% employed full-time)		130 (47.3)
Phone plan includes text messaging		266 (96.1)
Type of text messaging plan (% unlimited)		197 (71.2)

Personal Goals

Participants generated a variety of personal goals to refer to while choosing their preferred messages. In total, 137 out of 277 participants (49.5%) generated personal goals related to physical health and well-being. Within this broad category, 52 out of 137 participants (38.0%) generated fitness goals (eg, "I want to go to the gym more often") and 44 out of 137 participants (32.1%) generated weight loss goals. Also, 103 out of 277 (37.2%) participants generated personal goals related to competence and mastery. Within this category, 33 of 103 participants (29.9%) generated financial goals (eg, "I want to save more money this year"), 25 of 103 participants (24.3%) generated professional goals (eg, "I want to advance in my company"), and 18 of 103 participants (17.4%) generated personal mastery goals (eg, "I want to build my own house"). In addition, 33 out of 277 participants (11.9%) generated goals related to personal fulfillment (eg, "I want to communicate more effectively with my spouse" or "I want to have more fun").

Message Preferences

Results of messaging preferences are presented in Table 3. For the intents and purposes of this paper, we define "clear preference" as a preference of 75% or more of respondents for one message type within a grouping. There were clear preferences for about half of the groupings, with more than 90% of respondents selecting messages that did not include textese, a sad emoticon, incorrect grammar, or an external locus of control in those four groupings. There were also strong global preferences for messages with a benefit-oriented or active tone and for polite messages. For many of the groupings, no clear preferences were found for the entire sample.

Table 3. Message grouping preferences[a] (n=277).

Message Type: Greater Preference	%	Message Type: Lesser Preference	%	n[c]
Smiley emoticon	97.6	Sad emoticon	2.4	213[b]
Correct grammar	96.7	Grammatical errors	3.3	211[b]
Non-textese	95.8	Textese	4.2	216[b]
Locus of control: intrinsic	93.5	Locus of control: extrinsic	6.5	46[b]
Benefit-oriented	89.2	Consequence-oriented	10.8	195[b]
Polite	86.5	Impolite	13.5	208[b]
Nonaggression	82.9	Aggression	17.1	269[b]
Direction	82.3	Passive	17.7	211[b]
Statement	82.0	Question	18.0	245[b]
No Humor	77.9	Humor	22.1	188[b]
Male quote	71.9	Female quote	28.1	146[b]
"I" statement	66.7	"We" statement	33.3	264[b]
Single punctuation	64.9	Multiple punctuation	35.1	174[b]
"You" statement	62.9	"We" statement	37.1	272[b]
Uncited	62.9	Cited	37.1	272[b]
Nondirective	61.0	Command	39.0	272[b]
Coaching	57.4	Uncoached direction	42.6	61
Literal	56.6	Metaphorical	43.4	272[b]
Smiley emoticon	53.6	No emoticon	46.4	274
CAPS (capitalization) emphasis	53.1	No visible emphasis	46.9	213
General goal	52.6	Implementation intention	47.4	57
Short	51.1	Long	48.9	272

[a] See Table 1 for a definition and example of each dyad.

[b] Indicates that a message preference is not the result of chance using a non-parametric binomial test to ensure that the difference between groups was greater than a 50% chance (*P*<.05).

[c] The n applies to both message types.

Message Revision

Because this was an intervention development study, we also created several messages in which we manipulated specific language components from sample to sample. Slightly altering the wording of the passive message within Directive vs Passive dyad #3 from "Every time you feel down, try to change your thoughts to something positive about change" to "Every time you feel down, you might want to try to change your thoughts to something positive about change" resulted in an increase in participants' overall preferences for the directive message: 127 out of 208 participants (61.1%) preferred the directive message prior to the dyad's change, but 51 out of 57 participants (89.5%) preferred the directive message post-change. Within a message grouping that examined preferences for Statistics vs Anecdotes, changing the statistic within message dyad #3 from "93% of people who monitor their food intake reduce their calorie intake" to "44% of people who monitor their food intake reduce their calorie intake" caused overall preferences for the statistic to diminish; while 156 out of 212 participants (73.6%) preferred the message containing a statistic in the first version, only 37 out of 60 participants (61.7%) preferred it after the statistic was changed. Conversely, changing the statistics in message dyad #4 from "People who report doing nice things for other people are 44% happier than those who do not" to "People who report doing nice things for other people are 93% happier than those who do not" caused the overall preferences for the statistic to increase: 97 out of 210 participants (46.2%) preferred the message containing a statistic prior to its change, while 37 out of 57 participants (64.9%) preferred the message after the statistic was changed.

Message Preference Moderators

We assessed differences in preferences based on several demographic variables, including gender, age, and education. A significant difference existed between male and female participants' preferences for messages in only one message grouping, with female participants being more likely than male participants to prefer correct grammar to incorrect grammar (χ^2_{212}=5.334, *P*=.021; male=93.5%, 86/92; female=99.2%, 120/121). Similarly, the only significant difference that existed between older and younger individuals' preferences was that older individuals were more likely than younger individuals to prefer "you" statements to "we" statements (χ^2_{271}=7.669, *P*=.006; over 40 years=76.4%, 55/72; 40 years or under=58.0%, 116/200).

Significant differences in preference existed between participants with different levels of education for several message groupings as well. Participants with less than a college degree were more likely than participants with a college degree or greater level of education to prefer directions to suggestions (χ^2_{210}=6.061, *P*=.014; no college=97.9%, 95/97; college=89.5%, 102/114), short messages to long messages (χ^2_{267}=3.759, *P*=.053; no college=55.1%, 70/127; college=43.3%, 61/141), and messages that included smiley emoticons to messages that contained no emoticons (χ^2_{269}=3.569, *P*=.059; no college=59.4%, 76/128; college=47.9%, 68/142). We ran a multiple logistic regression with significant moderators as the independent variables and education as the dependent variable. When controlling for all variables, education only moderated the preferences for the direction vs passive suggestion grouping (Wald statistic $_{3,205}$ =5.26, *P*=.022). There were no differences in message preferences based on employment status.

We also assessed differences in preference based upon personality or trait variables. Participants who reported being generally sad were significantly more likely than participants who reported being generally happy to prefer commands to nondirective general statements (χ^2_{267}=4.037, *P*=.045; sad=50.8%, 30/59; happy=36.4%, 76/209), literal language to metaphors (χ^2_{267}=6.508, *P*=.011; sad=70.7%, 40/58; happy=51.9%, 109/210), non-polite messages to polite messages (χ^2_{204}=3.907, *P*=.048; sad=22.7%, 10/44; happy=11.2%, 18/161), and loss-framed to gain-framed messages (χ^2_{190}=4.193, *P*=.041; sad=20.5%, 9/40; happy=8.6%, 13/151). There was also a trend for this group to prefer "I" statements to "We" statements (χ^2_{259}=3.136, *P*=.077; sad=76.8%, 43/56; happy=64.2%, 131/204). Multiple logistic regression with significant moderators revealed that only the preferences of "I" statements to "We" statements (Wald statistic $_{4,179}$ =4.74, *P*=.029) and commands to nondirective general statements (Wald statistic $_{4,179}$ =8.29, *P*=.004) remained significant.

Despite the fact that participants reported radically different goals, the only differences that existed between participants with different higher order goals were that those with personal fulfillment goals were significantly more likely to prefer consequence messages than either those with physical health and well-being or competence and mastery goals (χ^2_{189} =6.829, *P*=.033; personal fulfillment=24.0%, 6/25; physical health and well-being=10.8%; 10/93; competence and mastery=5.5%, 4/73). However, a heavy preference for benefit-based messaging existed across all three groups. The relationship between goals and messaging preferences will be discussed at length in a future paper.

Finally, we assessed differences in preferences based on three process rulers: the participants' perceived benefits of changing, confidence about their ability to change, and perceived importance of changing. A preference for coaching messages was significantly associated with perceiving greater benefits of change ($F_{1,31}$=4.33, *P*=.047). A preference for caps (capitalization) emphasis was significantly associated with perceiving greater benefits of change ($F_{1,110}$=4.719, *P*=.032), higher confidence in one's ability to change ($F_{1,197}$=6.732, *P*=.012), and perceiving greater importance of changing ($F_{1,200}$=9.325, *P*=.003). When entered into a logistic regression and with the benefits of changing variable removed due to a small sample size, only goal importance remained significant (Wald statistic $_{2,197}$ =6.78, *P*=.009) while goal confidence trended towards significance (Wald statistic $_{2,197}$ =3.43, *P*=.064).

Discussion

Principal Findings

To our knowledge, this is the first study to quantitatively examine messaging preferences for a range of message types to help guide text-based mobile intervention development. Results of this study indicate that there are clear user preferences for certain types of message characteristics over others, underscoring the importance of attention to message structure, linguistic content, and overall tone in the development of messages for goal-directed behaviors. This is particularly true of accurate spelling and grammar, as well as messages that emphasize the positive over the negative. While there has been little quantitative research on this topic, the findings of the present study are generally reflective of past qualitative research on messaging development, and are further supported by the Centers for Disease Control and Prevention's *Guide to Writing for Social Media* [27]. For guidelines on writing messages for goal-directed behaviors based on this research and other sources, please see Multimedia Appendix 1.

Spelling, Punctuation, and Grammar

Participants indicated an overwhelming preference for messages that were accurately spelled and grammatically correct over messages that included *textese* or contained grammatical errors. *Textese* can be more difficult to process than properly spelled words and phrases [28]. It is possible that participants' overwhelming preference for proper spelling over *textese* is due to the fact that *textese* can impede comprehension and therefore reduce message receptivity. Another explanation for this finding that has been suggested in qualitative studies is that the inclusion of *textese* or the accidental inclusion of spelling and grammatical errors threatens the source credibility of messages that are designed to help users achieve a goal [8,29]. This may also explain participants' preferences for serious messages over messages that attempted to be humorous.

The third and fourth message groupings to examine message syntax compared messages with single punctuation (eg, "." or "!") to messages with multiple punctuation marks (eg, "…" or "!!!") and capitalization of a whole word or phrase to no capitalization (eg, "When it comes to the negative consequences of a bad habit, you are NOT the exception"). Multiple punctuation marks or capitalization can be utilized to add emphasis or to create a pause, and thus constitute visual substitutions for verbal cues. Participants' preferences for single over multiple punctuation marks was much less pronounced than their preferences for proper spelling and grammar, and there was no clear preference between messages with caps emphasis versus messages with no visible emphasis. However, participants who reported that meeting their goal was very important and would benefit them immensely were more likely to prefer messages with some or all capitalized words for emphasis, and there was a trend toward a significant preference for multiple punctuation marks in the high benefits of change group. This emphasizes that understanding the end user's state is a crucial component of intervention development.

Emoticons

We examined variations in preferences for two different message groupings that contained emoticons: one that compared messages with a smiley face to identical messages with a sad face, and another that compared messages with a smiley to messages with no emoticons. Of all of the message groupings we examined, preferences were strongest in the smiley versus sad emoticon message grouping, with participants vastly preferring messages with the smiley. This finding resonates with past research that suggests that users vastly prefer positive messages to negative messages, as does our finding that participants generally preferred benefit-oriented to consequence-oriented messages. By contrast, no clear preference existed for messages that contained a smiley versus messages that contained no emoticon. It is possible that some participants found the inclusion of an emoticon too informal within messages designed to help users achieve a personal goal, while others perceived the inclusion of a smiley face as encouraging or rewarding when compared to a message with no emoticon. While virtually all participants seem to prefer a positive to a negative image, the fact that preferences within the two emoticon groupings varied so extremely suggests that the inclusion of an emoticon can communicate very different things depending on the context.

In concordance with previous literature suggesting that visual cues are more effective than text for those with lower need for cognition [15,30], we found that individuals with less education were more likely to prefer the inclusion of an emoticon than those with more education. Understanding who may be more receptive to emoticons [31] is an important line of research due to the frequency in which they are integrated into existing health messaging interventions. When taken together, these results underscore the need to tailor communication patterns to individual differences to obtain maximum engagement in goal-directed interventions.

Sentence Type: Declarative, Interrogative, and Imperative

The clear user preferences for statements over questions have particularly interesting implications, as self-evaluative questions are often used in order to integrate motivational interviewing techniques into messaging interventions. Based on our findings, the inclusion of such messages without a fuller understanding of the preferences of the end-user requires some reconsideration. For example, will an individual who drinks too much be motivated to contemplate and evaluate the consequences of drinking simply because a text message asks him or her to do so? Further, questions that do not require interactivity may be disregarded by the individual because they will receive no feedback on their response. On the other hand, Muench and colleagues [23] found that individuals who are in the process of thinking about engaging in problem behaviors requested self-evaluative messages, suggesting that participants may be more receptive to such questions at specific times or stages of change.

There were clear preferences for more directive language over passive or suggestive language. Moderator analysis revealed that the preference for directive language was especially pronounced in older adults and individuals with more education. This could be a result of several factors, but may simply indicate that these users have been taught to avoid passive or suggestive language when communicating. While participants' overall preference for directive over passive messages was clear, they were more averse to commands for immediate action (eg, "Do x right now…"), indicating that while individuals may want instruction, they may not want to feel commanded to behave a certain way in the moment.

Sentence Content, Grammatical Person, and Length

As mentioned above, there was also a general preference for messages that did not include humor. It could be that humor minimizes one's struggle to achieve a goal and should be used sparingly and possibly only after an alliance is built. Messages that were presented in first person singular ("I" statements) and second person ("You" statements) were preferred to messages that were presented in first person plural ("We" statements). This may indicate that participants generally prefer to be identified as individuals as opposed to one of a number of people, and prefer to identify the message originator as an individual as well.

We found no clear preferences for several other types of messages within groupings, including short vs long messages. This grouping is of particular interest because intervention developers have often been encouraged to break down messages to their smallest component pieces [32]. By contrast, our findings suggest that there are instances in which shortening a message can constitute a sacrifice in its readability, meaning, or cohesiveness, which should be avoided where possible. While there was a general preference for uncited statements over cited statements, we found that increasing the percentage of the effect in a statistic (eg, 44% to 93%) resulted in more pronounced preferences for the statistic. This finding corresponds well to the research on conformity and the power of social norms to increase the persuasiveness of messages. It is also possible that there are significant differences in how different outcomes (eg, happiness vs mortality) might significantly alter the persuasiveness of differing statistics, and this should be explored in greater detail.

Group Differences

Differential preference analysis was designed to help distinguish preferences among different groups. In our case, overall analyses did not reveal dramatic shifts in preferences, but rather subtle differences between certain groups on certain variables. For example, while younger participants were significantly more likely than older participants to prefer "We" messages to "You" messages, the two groups differed by only 14%. In fact, few moderators shifted one group's overall preferences for a dyad from one message type to the other. However, the differences reported are significant and future research should examine these subtle variations in message preferences based on these differences. For example, Muench and colleagues [23] found that in a substance abuse sample, individuals generally preferred benefit-oriented messaging, but that this was more pronounced with individuals who reported higher perceived benefits of changing. Because this finding is similar to the general health tailoring literature [33], examining moderators of preferences may be an efficient way to develop tailored interventions.

Interestingly, some individuals preferred more negative messaging. As there is ample evidence that aggression or a demeaning tone decreases long-term adherence (as opposed to constructive negative feedback or consequence-oriented messaging, which can be useful), this finding highlights the downside of preference research in guiding intervention design. Communications that contain negative components like shaming or punishment are contraindicated in interventions to promote long-term behavior change [34], even if some people claim to prefer these messages. So why do some individuals prefer impolite or aggressive messages and how can we identify this group? For example, post-hoc analysis revealed that the small group who preferred the sad emoticon also preferred more aggressive messaging. Similarly, we found that individuals who endorsed being generally sad were significantly more likely than those who reported being happy to choose consequence-oriented and literal messages, and there was a trend for choosing "I" messages over "We" messages. In cases such as this, preferences might simply be an assessment tool to understand the individual, and tailoring to these types of preferences may not improve outcomes. At the same time, it is possible that sending messages congruent with preferences can increase engagement in an intervention. Then as the user becomes more engaged, there can be a slow and subtle shift to "healthier" messaging. For example, an intervention can send individuals with depression messages with "I" statements early in the change process and then slowly shift to "We" statements as the intervention progresses. The general tailoring literature further indicates that engagement in digital interventions is one of the best predictors of outcome [35].

It was surprising to us that there were few differences in preferences between those with different scores on process rulers, with the only differences being that those with higher confidence, importance, and benefits to changing preferred caps emphasis in messages to no emphasis and those who saw greater benefits of meeting their goal preferred coaching messages when compared to those who perceived few benefits in meeting their goal. Both messaging types are designed to illicit some sort of emotion in the individual and it could be that adding a positive emotional emphasis is useful for individuals who see greater benefits or importance for change or have greater confidence in their ability to change. While understanding these processes certainly has important implications for message development, it was more striking to us that there were so few differences and results should be interpreted with caution.

Limitations

Despite the promise of this line of research, there are limitations. The most salient is that preferences do not necessarily translate into improved outcomes [36] and may, in fact, reflect the underlying negative schemas of an individual, such as a tendency to self-shame or diminish self-efficacy. Therefore, regardless of preferences, certain precautions should be taken when individualizing messages, such as avoiding aggressive language or sad emoticons, whereas more leeway can be given with messages that are neutral or moderately contraindicated (eg, fostering an external locus of control) if it improves engagement. Another limitation is that we did not send actual mobile messages, but asked for preferences in an online survey. It is possible that viewing a message on a computer screen as opposed to in an actual text or mobile alert may yield different preferences. This is particularly true of preferences for variables like *textese*, which are most commonly seen in mobile communications. Nevertheless, our results are congruent with much of the general health communications literature in terms of outcomes [37], indicating that preferences research is a useful preliminary step in intervention development. Results of our moderator analyses should be interpreted with reservation as we did multiple analyses, inflating the possibility of a Type-II error. However, we performed logistic regression with significant variables to look at the unique contributions of each moderator to messaging preferences.

There were population limitations as well. Namely, we did not restrict the availability of the survey only to people who might be the target of a health intervention, but left it open to a wider population with a broad range of goals—some of which were completely unrelated to health. It is possible that this wider population may have different message preferences than a sample of people who are struggling specifically with a health or mental health problem. Because some of the message groupings were added during later cohorts, there were smaller sample sizes for these preference findings, reducing the strength of the effects. Therefore, future studies should replicate these findings with larger samples. Finally, we limited this analysis to US populations. As a larger part of the study, we are comparing US and Indian populations on messaging preferences, as there is good evidence that linguistic styles and communications differ dramatically between cultures [38]. Despite some limitations, using bottom-up participatory strategies can help us design interventions that account for the client's preferences within a larger theoretical framework. Adding quantitative preference testing yields similar results as focus groups with significantly lower burden, and can be iterated and modified quickly to test subtle variations in our models of care.

Conclusions and Future Research

When taken together, understanding preferences for intervention presentation may improve engagement, regardless of the content or theory upon which an intervention is based [39-41], and is therefore a logical extension of mobile messaging research. While we do not know the impact of differentially structuring messages on actual outcomes, they may increase engagement in message-based interventions. For example, a recent study revealed that while emoticons did increase users' enjoyment of the texting interaction, the perceived enjoyment of emoticons had no effect on the perceived usefulness of the message [42]. Once we understand these global messaging preferences, we can begin to further examine preference moderators such as those reported in the health tailoring literature [11,12]. Moreover, research should examine user preferences for actual text messages using ecological momentary assessment and response rate feedback. This research will provide investigators with a better understanding of preferences in real-world contexts. Future research can also focus on receptivity as it relates to specific periods in the change process (eg, an emoticon when one is meeting one's goal compared to an emoticon when one is not). Muench and colleagues [23] revealed that individuals generate different messages for different critical points in the change process (eg, when someone is at risk for relapse vs someone who has already relapsed), underscoring that in dynamically tailored interventions, messages must contain appropriate content at appropriate moments, and this may apply to message structure as well. Using consumer research methods such as rapid iterative design, in which surveys are republished multiple times with slight changes to content, can expedite intervention development research. Because just-in-time mobile interventions are in their infancy, this line of research can help guide researchers to test proposed intervention components prior to conducting larger scale trials. As a preliminary step, we developed general guidelines for writing messages for goal-directed behavioral messaging interventions. We purposely kept it short and presented the most robust findings. However, this should be considered preliminary until more research can be done.

Acknowledgments

This study was supported with funding from the National Institute on Alcohol Abuse and Alcoholism (R34 AA021502-01).

Conflicts of Interest

Dr Muench consults with several mobile health companies on the development of mobile messaging platforms for health concerns. There were no other conflicts of interest.

Multimedia Appendix 1

Tips for writing intervention messages based on user preferences.

[PDF File (Adobe PDF File), 168KB - jmir_v16i2e14_app1.pdf]

References

1. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
2. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010 Apr;32(1):56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
3. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]

4. 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]
5. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
6. Kraft P, Drozd F, Olsen E. ePsychology: Designing theory-based health promotion interventions. Communications of the Association for Information Systems 2009;24(1):399-426.
7. Bickmore T, Schulman D, Yin L. Maintaining engagement in long-term interventions with relational agents. Appl Artif Intell 2010 Jul 1;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
8. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: Key issues, process model, and system features. Communications of the Association for Information Systems 2009;24(1):485-500.
9. Bickmore T, Mauer D, Crespo F, Brown T. Persuasion, task interruption and health regimen adherence. Persuasive Technology: Springer; 2007:1-11.
10. Goetz J, Kiesler S, Powers A. Matching robot appearance and behavior to tasks to improve human-robot cooperation. 2003 Presented at: 12th IEEE International Conference on Data Mining; 2003; Brussels, Belgium. [doi: 10.1109/ROMAN.2003.1251796]
11. Noar SM, Harrington NG, Aldrich RS. The role of message tailoring in the development of persuasive health communication messages. In: Beck CS, editor. Communication Yearbook 33. New York: Lawrence Erlbaum Associates, Inc; 2009:73-133.
12. Dijkstra A. The Psychology of Tailoring-Ingredients in Computer-Tailored Persuasion. Social Pers Psych Compass 2008 Mar;2(2):765-784. [doi: 10.1111/j.1751-9004.2008.00081.x]
13. Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: the case for tailoring print materials. Ann Behav Med 1999;21(4):276-283. [Medline: 10721433]
14. Cacioppo JT, Petty RE, Feinstein JA, Jarvis WBG. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. Psychological Bulletin 1996;119(2):197-253. [doi: 10.1037/0033-2909.119.2.197]
15. Williams-Piehota P, Schneider TR, Pizarro J, Mowad L, Salovey P. Matching health messages to information-processing styles: need for cognition and mammography utilization. Health Commun 2003;15(4):375-392. [doi: 10.1207/S15327027HC1504_01] [Medline: 14527864]
16. Walther JB, D'Addario KP. The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication. Social Science Computer Review 2001 Aug 01;19(3):324-347. [doi: 10.1177/089443930101900307]
17. Woolford SJ, Barr KL, Derry HA, Jepson CM, Clark SJ, Strecher VJ, et al. OMG do not say LOL: obese adolescents' perspectives on the content of text messages to enhance weight loss efforts. Obesity (Silver Spring) 2011 Dec;19(12):2382-2387 [FREE Full text] [doi: 10.1038/oby.2011.266] [Medline: 21869762]
18. Gold J, Lim MS, Hocking JS, Keogh LA, Spelman T, Hellard ME. Determining the impact of text messaging for sexual health promotion to young people. Sex Transm Dis 2011 Apr;38(4):247-252. [doi: 10.1097/OLQ.0b013e3181f68d7b] [Medline: 20966830]
19. Sakai Y. The role of readability in effective health communication: an experiment using a Japanese health information text on chronic suppurative otitis media. Health Information & Libraries Journal 2013:220-231. [doi: 10.1111/hir.12022]
20. Leroy G, Helmreich S, Cowie JR. The influence of text characteristics on perceived and actual difficulty of health information. Int J Med Inform 2010 Jun;79(6):438-449. [doi: 10.1016/j.ijmedinf.2010.02.002] [Medline: 20202895]
21. Kraichy D, Chapman DS. Tailoring Web-Based Recruiting Messages: Individual Differences in the Persuasiveness of Affective and Cognitive Messages. J Bus Psychol 2013 Jun 2:1. [doi: 10.1007/s10869-013-9311-z]
22. Metzler CW, Sanders MR, Rusby JC, Crowley RN. Using consumer preference information to increase the reach and impact of media-based parenting interventions in a public health approach to parenting support. Behav Ther 2012 Jun;43(2):257-270 [FREE Full text] [doi: 10.1016/j.beth.2011.05.004] [Medline: 22440064]
23. Muench F, Weiss RA, Kuerbis A, Morgenstern J. Developing a theory driven text messaging intervention for addiction care with user driven content. Psychol Addict Behav 2013 Mar;27(1):315-321. [doi: 10.1037/a0029963] [Medline: 22963375]
24. Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods 2012 Mar;44(1):1-23. [doi: 10.3758/s13428-011-0124-6] [Medline: 21717266]
25. 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]
26. Sobell MB, Sobell LC. Problem Drinkers: Guided self-change treatment. New York: Guilford Press; 1996.
27. O'Flavahan L, Goulet A. www.CDC.gov. 2012. CDC's Guide to Writing for Social Media URL: http://www.cdc.gov/socialmedia/tools/guidelines/pdf/guidetowritingforsocialmedia.pdf [accessed 2013-12-20] [WebCite Cache ID 6M0fmCATd]
28. Kemp N. Texting versus txtng: reading and writing text messages, and links with other linguistic skills. Writing Systems Research 2010 Jan;2(1):53-71. [doi: 10.1093/wsr/wsq002]
29. Liu Z. Perceptions of credibility of scholarly information on the web. Information Processing & Management 2004;40(6):1027-1038.
30. Petty RE, Briñol P, Loersch C, McCaslin MJ. The Need for Cognition. In: Leary MR, Hoyle RH, editors. Handbook of Individual Differences in Social Behavior. New York: The Guilford Press; 2009:318-329.

31. Xu L, Yi C, Xu Y. Emotional Expression Online: the Impact of Task, Relationship and Personality Perception on Emoticon Usage in Instant Messenger. 2007 Presented at: Pacific Asia Conference on Information Systems; July; Auckland, New Zealand.
32. Zarcadoolas C. The simplicity complex: exploring simplified health messages in a complex world. Health Promot Int 2011 Sep;26(3):338-350 [FREE Full text] [doi: 10.1093/heapro/daq075] [Medline: 21149317]
33. Rothman AJ, Updegraff JA. Specifying when and how gain-and loss-framed messages motivate healthy behavior: An integrated approach. In: Keren G, editor. Perspectives on Framing. New York: Psychology Press; 2010:257-277.
34. Bayer R. Stigma and the ethics of public health: not can we but should we. Soc Sci Med 2008 Aug;67(3):463-472. [doi: 10.1016/j.socscimed.2008.03.017] [Medline: 18502551]
35. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
36. Lawton J, Rankin D, Elliott J, United Kingdom National Institute for Health Research Dose Adjustment for Normal Eating (DAFNE) Study Group. Is consulting patients about their health service preferences a useful exercise? Qual Health Res 2013 Jul;23(7):876-886. [doi: 10.1177/1049732313485161] [Medline: 23618684]
37. Latimer AE, Rivers SE, Rench TA, Katulak NA, Hicks A, Hodorowski JK, et al. A Field Experiment Testing the Utility of Regulatory Fit Messages for Promoting Physical Activity. J Exp Soc Psychol 2008 May;44(3):826-832 [FREE Full text] [doi: 10.1016/j.jesp.2007.07.013] [Medline: 19543443]
38. Calderón JL, Beltran B, Beltran B. Culture and linguistics: neglected variables in the health communication equation. Am J Med Qual 2005;20(4):179-181. [doi: 10.1177/1062860605276033] [Medline: 16020674]
39. 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. American Journal of Health Promotion 1998;13(1):12-18.
40. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
41. Kaulio MA. Customer, consumer and user involvement in product development: A framework and a review of selected methods. Total Quality Management 1998;9(1):141-149.
42. Lim SH, Kim D, Watts S. An Empirical Study of the Usage Performance of Mobile Emoticons-Applying to the Five Construct Model by Huang. Journal of Information Technology Applications & Management 2011;18(4):21-40.

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
HIT: human intelligence task
MTurk: Amazon Mechanical Turk
SMS: short message service

Edited by D Mohr; submitted 13.09.13; peer-reviewed by S Schueller, A Aguilera, A Morgan, A Ramachandran; comments to author 28.09.13; revised version received 19.11.13; accepted 21.11.13; published 05.02.14.

Please cite as:
Muench F, van Stolk-Cooke K, Morgenstern J, Kuerbis AN, Markle K
Understanding Messaging Preferences to Inform Development of Mobile Goal-Directed Behavioral Interventions
J Med Internet Res 2014;16(2):e14
URL: http://www.jmir.org/2014/2/e14/
doi:10.2196/jmir.2945
PMID:24500775

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