Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 17 (2015), Issue 4 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Original Papers

Viewpoint

Review

Corrigenda and Addenda

Review

The Use of Behavior Change Theory in Internet-Based Asthma Self-Management Interventions: A Systematic Review

Mustafa Al-Durra[1,2], BSc, MSc; Monika-Bianca Torio[1,3], BASc; Joseph A Cafazzo[1,2,4], PhD, PEng

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

[2]Institute of Health Policy, Management and Evaluation, The Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]Faculty of Applied Science & Engineering, University of Toronto, Toronto, ON, Canada

[4]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Mustafa Al-Durra, BSc, MSc
Centre for Global eHealth Innovation, Techna Institute, University Health Network
Toronto General Hospital, R Fraser Elliott Building, 4th Floor
190 Elizabeth Street
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 4800 ext 4765
Fax: 1 416 340 3595
Email: maldurra@ehealthinnovation.org

Abstract

Background: The high prevalence rate of asthma represents a major societal burden. Advancements in information technology continue to affect the delivery of patient care in all areas of medicine. Internet-based solutions, social media, and mobile technology could address some of the problems associated with increasing asthma prevalence.

Objective: This review evaluates Internet-based asthma interventions that were published between 2004 and October 2014 with respect to the use of behavioral change theoretical frameworks, applied clinical guidelines, and assessment tools.

Methods: The search term (Asthma AND [Online or Internet or Mobile or Application or eHealth or App]) was applied to six bibliographic databases (Ovid MEDLINE, PubMed, BioMed Central, ProQuest Computing, Web of Knowledge, and ACM Digital Library) including only English-language articles published between 2004 and October 2014. In total, 3932 articles matched the priori search terms and were reviewed by the primary reviewer based on their titles, index terms, and abstracts. The matching articles were then screened by the primary reviewer for inclusion or exclusion based on their abstract, study type, and intervention objectives with respect to the full set of priori inclusion and exclusion criteria; 331 duplicates were identified and removed. A total of 85 articles were included for in-depth review and the remaining 3516 articles were excluded. The primary and secondary reviewer independently reviewed the complete content of the 85 included articles to identify the applied behavioral change theories, clinical guidelines, and assessment tools. Findings and any disagreement between reviewers were resolved by in-depth discussion and through a consolidation process for each of the included articles.

Results: The reviewers identified 17 out of 85 interventions (20%) where at least one model, framework, and/or construct of a behavioral change theory were applied. The review identified six clinical guidelines that were applied across 30 of the 85 interventions (35%) as well as a total of 21 assessment tools that were applied across 32 of the 85 interventions (38%).

Conclusions: The findings of this literature review indicate that the majority of published Internet-based interventions do not use any documented behavioral change theory, clinical guidelines, and/or assessment tools to inform their design. Further, it was found that the application of clinical guidelines and assessment tools were more salient across the reviewed interventions. A consequence, as such, is that many Internet-based asthma interventions are designed in an ad hoc manner, without the use of any notable evidence-based theoretical frameworks, clinical guidelines, and/or assessment tools.

(J Med Internet Res 2015;17(4):e89) doi:10.2196/jmir.4110

KEYWORDS

asthma; self-care; self-management; eHealth, mHealth; mobile health; telehealth; telemedicine; Internet; review

Introduction

Asthma is a common chronic inflammatory disease of the airways with symptoms including cough, breathlessness, and wheezing. According to World Health Organization (WHO) estimates, there are some 235 million people in the world currently suffering from asthma. The WHO also estimates that asthma is the most common non-communicable disease among children [1,2].

Combined with the aging population trend and increasing cost of health care services, the high prevalence rate of asthma represents a major societal burden as well as a substantial challenge to the traditional models of health care providers, patients, and their families. A number of cost analysis studies have reported that the annual economic cost of asthma due to direct medical costs from hospital stays, as well as indirect costs from lost school and workdays, amounted to more than US $56 billion in the United States in 2007, CAN $1.8 billion in Ontario, Canada in 2011, and €19.3 billion in European adult populations in 2010 [3-5].

Advancements in the field of information technology continue to change patient care in all areas of medicine. Internet-based solutions, social media, and mobile technology could help to mitigate some of the problems associated with the increasing asthma prevalence [6].

The National Heart, Lung, and Blood Institute Expert Panel Report 3 (NHLBI EP3) Asthma Guidelines suggest that there is a potential use for information technologies to provide patients with skills to control their asthma and improve outcomes [7,8].

None of the existing literature reviews focused on evaluating Internet-based asthma interventions with respect to the evidence base around the behavioral change theoretical frameworks, applied clinical guidelines, and assessment tools.

The primary objective of this literature review was to identify and evaluate Internet-based asthma interventions that were published between 2004 and October 2014 with respect to the use of the behavioral change, self-care, and self-management theoretical frameworks as well as the application of clinical guidelines and assessment tools.

Methods

Research Questions

We established the following primary research question: What is the use of behavioral change, self-care, and self-management theoretical frameworks within the context of Internet-based asthma interventions?

Our secondary research question was: What is the use of clinical guidelines and assessment tools within the context of Internet-based asthma interventions?

Inclusion Criteria

The review included all asthma-related Internet-based interventions, such as Internet-based applications, electronic diary solutions, mobile apps, and/or any other kind of computer-based applications with the focus on patient-centric Internet-based applications as well as provider-to-patient applications.

The bibliographic databases search included relevant studies and interventions that were published between 2004 and October 2014 and was limited to literature published in the English language**.**

Exclusion Criteria

The review excluded any electronic record management systems that are provider-centric and used to organize patient visits at the clinic and/or hospital settings such as electronic medical records (EMR), electronic health records (EHR), and hospital information systems (HIS). Also, telemedicine interventions that merely leveraged the conventional wired or wireless telephone technology as a medium to facilitate a verbal communication and/or short message service (SMS) between patients and their providers were excluded. The review excluded any educational-only studies that utilized Web-based resources, such as social media, decision support tools, and wikis, for the sole purpose of providing educational content for asthma patients, caregivers, and/or providers.

The bibliographic databases search excluded studies whose main objective was to design, develop, and assess eHealth tools, such as Web, Internet, and mobile apps, without providing critical analysis of their impact and contribution within a given asthma intervention context.

Search Strategy

Overview

The search term was applied to six bibliographic databases (Ovid MEDLINE, PubMed, BioMed Central, ProQuest Computing, Web of Knowledge, and ACM Digital Library) including only English articles published between 2004 and October 2014. The search was conducted in the following steps.

Search Term

We limited the search to English-language articles published between 2004 and October 2014. The search term was:

- [Asthma] AND
- [English language and year="2004 -Current"] AND
- [Online or Internet or Mobile or Application or eHealth or App]

Step I — Abstract Evaluation

The primary reviewer evaluated the abstracts, titles, and index terms of all matching articles in the bibliographic databases where the search term was applied. Based on this preliminary review, all relevant articles were listed for potential inclusion.

Step II — Screening for Inclusion

In this step, the primary reviewer evaluated relevant articles in the preliminary list for final inclusion or exclusion based on their abstract, study type, and intervention focus with respect to the full set of priori inclusion and exclusion criteria.

Step III — Removal of Duplicates

Internal and cross-database duplicates were identified and removed from all included articles from Step II. Duplicates within each database were first identified and removed. Cross-database duplicates were then identified and removed through a manual consolidation process.

Step IV — Independent Review

The complete published papers of all included articles were then reviewed, analyzed, and assessed thoroughly by two reviewers independently. The primary and secondary reviewers independently reviewed the complete content of the included articles to identify the applied theoretical frameworks, clinical guidelines, and assessment tools with the objective to answer the priori research questions. Findings and disagreements between the primary and secondary reviewers were resolved by in-depth discussions and through a consolidation process for each of the included articles.

Results

Overview

In total, and across all six bibliographic databases, 3932 articles matched the priori search terms and were reviewed by the primary reviewer based on their titles, index terms, and abstracts in Step I.

In Step II, 3516 articles were excluded by the primary reviewer on their abstract, study type, and intervention focus that met the priori exclusion criteria.

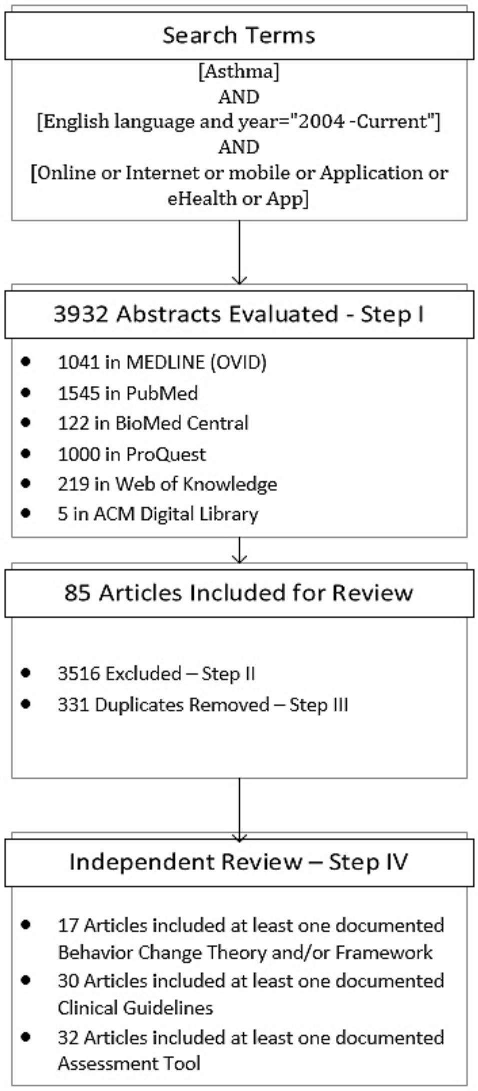
A total of 331 duplicates were identified and removed in Step III.

In the last step, the remaining 85 articles were included for independent and in-depth review by the two reviewers. Figure 1 depicts the search breakdown and results for all six bibliographic databases.

The majority of the reviewed studies and interventions reported the following key targeted behaviors [9-13]: (1) managing environmental triggers, (2) accessing asthma services, (3) medication adherence, (4) monitoring peak flow regularly by using portable meters, (5) keeping rescue inhaler accessible, and (6) smoke reduction or cessation.

The findings of the review results will be discussed in three different sections: Theoretical Frameworks, Clinical Guidelines and Assessment Tools, and Other Reviews.

Figure 1. Search results from the six bibliographic databases.



Theoretical Frameworks

Overview

The motive behind conducting this review was to answer the primary research question with respect to whether existing Internet-based interventions for asthma were founded on any behavioral-change theories. And if so, to what extent did these theoretical frameworks inform the design and evaluation of these interventions?

This review identified 17 out of 85 interventions (20%) where at least one model, framework, and/or construct of a behavioral change theory was applied. This implies that the majority of our reviewed interventions did not apply any documented behavioral change theory to inform the design of their interventions. This finding is consistent with what is reported in the literature. Theory-driven strategies for aiding individuals in changing or managing health behaviors are lacking [9].

As such, this review found that there are only a limited number of well-established behavioral change theories and models that were referenced and applied across multiple studies. In total, the reviewers were able to identify 10 behavioral change theories and models that were applied across multiple interventions versus 13 other theories and models that were only applied once within the context of a single study and/or intervention.

Table 1 provides a list of all applied theoretical frameworks and models that were identified across the 85 reviewed interventions.

In the following sections, the theoretical frameworks that were applied in more than three studies will be further analyzed and discussed.

Table 1. Applied theoretical frameworks and models of the 85 reviewed Internet-based asthma interventions.

Theoretical frameworks	Number of studies	Cited interventions
Gamification	4	[13-16]
Health Belief Model	4	[10-12,17]
Tailoring	4	[10-13]
Transtheoretical Model	4	[9,10,12,17]
Attribution Theory	3	[11,13,17]
Chronic Care Model	2	[18,19]
Motivational Interviewing	2	[11,17]
Self-Determination Theory	2	[17,20]
Social Cognitive Theory	2	[13,21]
Technology Acceptance Model	2	[22,23]
Biobehavioral Family Model (BFM)	1	[16]
Dual Processing Theory	1	[13]
Ecological Systems Theory	1	[24]
Instructional Theory	1	[13]
Intervention Mapping	1	[13]
Marlatt’s Theory of Relapse	1	[17]
Motivational Theory	1	[13]
Norma Engaging Multimedia Design (NEMD)	1	[16]
Peplau’s Theory of Interpersonal Relations	1	[22]
Social Learning Theory	1	[25]
Sociohistoric Theory	1	[13]
The eHealth Behavior Management Model	1	[9]
Theory of Planned Behavior	1	[9]
Watson’s Model of Caring	1	[22]

Gamification

In the past, computer and video games were perceived to be a waste of time and harmful in many aspects to those who play such games excessively, especially for the child and adolescent age groups [26,27]. Nevertheless, the advancement in audio-visual and telecommunication technologies has ignited a new era for today’s games. While the term “gamification” is still evolving, it could be defined as “the use of video game elements in non-gaming systems to improve user experience (UX) and user engagement” [28].

There is a growing body of evidence emphasizing the potential benefits of the social, health, and educational science behind computer games [29]. The potential application of computer games in the health domain was well addressed by the Games for Health projects. The project has defined a taxonomy to depict five main types of games used in health care: Preventative, Therapeutic, Assessment, Educational, and Informatics [27].

This review has shown that principles, concepts, and strategies of gamification were only applied in four studies. However, these four studies only targeted children up to 12 years old [13-16]. The reviewers could not cite any Internet-based asthma interventions employing gamification concepts for the adolescent or adults' population groups.

In one study conducted in 2013, an online peer support group for asthmatic children used an existing commercial networking website, Club Penguin, to help asthmatic children deal with difficult situations in an engaging manner [14].

Another two studies reported the success of an award-winning program called "Okay With Asthma", where an interactive digital story was developed and delivered online to support children with their asthma and psychosocial management strategies. This was done through leveraging and employing a novel behavioral model, the Biobehavioral Family Model [15,16].

The "Okay With Asthma" program successfully used the five factors (simulation interactivity, construct interactivity, immediacy, feedback, and goals) identified by the Norma Engaging Multimedia Design model to design its usability and feasibility testing approach [16,30].

> *Simulation interactivity describes the child's ability to 'become' a character in the story, whereas construct interactivity refers to the availability of activities for the child to create or build in the virtual world. Immediacy is the user's ability to observe all the actions and interactions that take place in the system. Children need feedback to show that their choices matter; without consequences, there would be no point in performing the actions. The model's final tenet is goal setting. Whether the goal is set extrinsically (by the game developer) or intrinsically (the child determining own goals), it is important for there to be goals to achieve. [31]*

As well, the Watch-Discover-Think-Act (WDTA) study [13] provided an applied example of how behavioral and motivation theories could be translated within the context of gamification. The WDTA program developed a game that walks through 18 real-life and four fantasy scenarios. The players, who are children with asthma, have to complete a set of tasks related to asthma self-management in order to progress across scenarios. Feedback is provided as a reinforcement of information for the children and their parents.

As depicted in Table 2, the reviewers were able to validate the translation steps of the behavioral change theoretical methods in the "Watch-Discover-Think-Act" [13] study against a number of other studies, such as the five factors of the Norma Engaging Multimedia Design model that were applied within the context of "Okay With Asthma" [16]. The correlation between the findings of those two different studies validates the impact and influence gamification theories and methods could have to increase patients' motivation, self-efficacy, and engagement level within the context of Internet-based asthma interventions.

Table 2. Correlation between the applied theoretical methods and factors of the two studies, "Watch-Discover-Think-Act" and "Okay With Asthma".

WDTA (Watch-Discover-Think-Act)	Okay With Asthma
Personalized Information	Simulation Interactivity
Fantasy Context + Multiple Modalities	Construct Interactivity
Learner Control	Immediacy
Reinforcement	Feedback
Goal Settings	Goals

Health Belief Model

The Health Belief Model dates back to the 1950s, initially developed by Hochbaum (1958) and Rosenstock (1960), and then extended by Kirscht and Becker in 1974 [32]. The Health Belief Model is a theoretical framework that attempts to study and predict the individual's health preferences and actions based on observed attitudes and personal beliefs. The model explains the individual's motivation to take a health care-related action based on the following factors:

> *(1) The existence of sufficient motivation (or health concern) to make health issues salient of relevant.*
>
> *(2) The belief that one is susceptible (vulnerable) to a serious health problem or to the sequelae of that illness or condition. This is often termed perceived threat.*
>
> *(3) The belief that following a particular health recommendation would be beneficial in reducing the perceived threat, and at a subjectively-acceptable cost. [33]*

The Health Belief Model was applied in the context of providing individualized messages and communication with patients to promote self-efficacy and better patient engagement [10-12,17].

The Puff City program that was evaluated in six Detroit high schools and reported in four different studies [10-12,17] identified and evaluated three core behaviors: namely, controller medication adherence, rescue inhaler availability, and smoking cessation/reduction. In the event of a negative change in any of the three core behaviors, theory-based health messages and information on asthma control were sent to the patients to sustain their self-efficacy and asthma self-regulation [12].

Tailoring

In the literature, "tailoring" is defined as "…assessment and provision of feedback based on information that is known or hypothesized to be most relevant for each individual participant of a program" [11,34,35], and "…any combination of information or change strategies intended to reach one specific person, based on characteristics that are unique to that person, related to the outcome of interest, and have been derived from an individual assessment" [10,35].

A number of studies and interventions have pointed out the significance of identifying the resistant groups at earlier stages of the intervention. These groups are less motivated to change their behavior and take ownership in managing their asthma. The objective is to use "tailoring" as a means to apply behavioral change theories, such as the Transtheoretical Model and the Health Belief Model, to motivate those subgroups and achieve positive changes in their behaviors with respect to their self-efficacy and asthma management [10-13]. The reviewed studies have utilized "tailoring" to customize the communication and education strategies with the targeted resistant subgroup based on their beliefs, attitude, and personalized information.

Transtheoretical Model

The Transtheoretical Model promotes individuals to change their behaviors for a healthier lifestyle.

> *The Transtheoretical Model is based on the premise that individuals are in one of five possible stages of change associated with a particular behavior. Precontemplation is the stage in which a person has no interest in changing the behavior. Contemplation is when a person would like to change the behavior someday but is not yet ready. Preparation is when a person is ready to make the change but needs assistance in moving that want into reality. The more active stages include Action and Maintenance. Those in Action have begun the behavior change process. Key to their success is moving the change to Maintenance, where change takes place over time. [9,36]*

The review identified three different studies where the concepts of the Transtheoretical Model were applied in the methods' design and patients' assessment through the stages of the change [9,10,12,17].

The Asthma Management Demonstration Project was developed to manage the following four asthma-related behaviors for asthma patients among employees and students of Western Michigan University: monitoring peak flow measurements, accessing asthma services, using prescription asthma medications properly, and managing environmental triggers [9]. Based on concepts of Transtheoretical Model, the project developed transactional questioning to stage their asthma patients according to their readiness to change their asthma-related behavior [9].

The Web-based Puff City program has also applied concepts of Transtheoretical Model to motivate their patients to change three core asthma-related behaviors: controller medication adherence, rescue inhaler availability, and smoking cessation/reduction [10,12,17].

Clinical Guidelines and Assessment Tools

Overview

In response to our secondary research question, we found that the application and employment of clinical guidelines and assessment tools were more abundant than theoretical frameworks across the reviewed interventions. The review identified 30 out of 85 interventions (35%) where at least one documented clinical guideline was applied. In total, there were six clinical guidelines applied across the 30 identified interventions as listed in Table 3.

As such, the review identified 32 out of 85 interventions (38%) where at least one documented assessment tool was applied. In total, there were 21 assessment tools applied across the 32 identified interventions as listed in Table 4.

This review found that many guidelines and assessment tools were broadly adopted by a relatively large number of interventions, for example, the National Asthma Education and Prevention Program referenced across 15 (of the 85) interventions [10-12,16-18,20-22,37-42], the Global Initiative for Asthma (GINA) guidelines referenced across 13 interventions [43-55], and the Asthma Control Questionnaire referenced across 11 interventions [19,20,44,49,56-62].

Table 3. Applied clinical guidelines of the 85 reviewed Internet-based asthma interventions.

Clinical guidelines	Number of studies	Cited interventions
National Asthma Education and Prevention Program (NAEPP)	15	[10-12,16-18,20-22,37-42]
Global Initiative for Asthma (GINA) Guidelines	13	[43-55]
British Guideline on the Management of Asthma	1	[54]
Canadian Asthma Consensus Guidelines (CACG)	1	[63]
International ERS/ATS Guidelines on Definition, Evaluation and Treatment of Severe Asthma	1	[54]
Standards for the Diagnosis and Treatment of Patients with COPD	1	[47]

Table 4. Applied assessment tools of the 85 reviewed Internet-based asthma interventions.

Assessment tools	Number of studies	Cited interventions
Asthma Control Questionnaire (ACQ)	11	[19,20,44,49,56-62]
Asthma Quality of Life Questionnaires (AQLQ)	9	[19,44,48,50,59,62,64-66]
Pediatric Asthma Quality of Life Questionnaire (PAQLQ)	8	[43,50,52,53,57,61,67,68]
Asthma Control Test (ACT)	8	[42,51-53,64,67-69]
International Survey of Asthma and Allergies in Childhood (ISAAC) questionnaire	4	[11,12,55,68]
Asthma Therapy Assessment Questionnaire (ATAQ)	3	[49,57,60]
Child Asthma Control Test (C-ACT)	3	[52,53,68]
Mini Asthma Quality of Life (Mini AQLQ)	3	[56,58,63]
Knowledge, Attitude and Self-Efficacy Asthma Questionnaire (KASE-AQ)	2	[58,60]
Pediatric Asthma Caregivers Quality of Life Questionnaire (PACQLQ)	2	[52,53]
The Asthma Life Quality Questionnaire (ALQ)	2	[56,64]
The Consumer Asthma Knowledge Questionnaire and inhalation technique with the standardized checklist of the Dutch Asthma Foundation	2	[57,59]
Air Quality Health Index (AQHI)	1	[63]
Allergic Rhinitis and its Impact on Asthma Questionnaire (ARIA)	1	[68]
Asthma Behavior Checklist (ABC)	1	[25]
Asthma Self-Regulatory Development Interview	1	[17]
Children’s Health Survey for Asthma (CHSA) by the American Academy of Pediatrics	1	[70]
Eyberg Child Behavior Inventory (ECBI)	1	[25]
Illness Management Survey (IMS)	1	[67]
Multidimensional Scale of Perceived Social Support	1	[11]
The Royal College of Physicians’ “Three Key Questions”	1	[71]

International Asthma Guidelines

With the aim to employ an evidence base to reduce asthma prevalence, morbidity, and mortality, the Global Initiative for Asthma (GINA) was launched in 1993 as a collaboration between the National Heart, Lung, and Blood Institute, the National Institutes of Health, and the World Health Organization [72]. The GINA guidelines were referenced in 13 of the 85 reviewed interventions [43-55].

Another example of internationally applied asthma guidelines is the International Survey of Asthma and Allergies in Childhood (ISAAC) questionnaire. Established in 1991, the ISAAC guidelines aimed to investigate asthma in the pediatric population as a measure to control the increasing conditions on the global scale [73]. Items from the ISAAC guidelines were included in the Lung Health Survey in four of the reviewed interventions [11,12,55,68].

National Asthma Guidelines

In response to the increasing asthma challenges in the United States, the National Asthma Education and Prevention Program (NAEPP) was initiated in March 1989 by the National Heart, Lung, and Blood Institute (NHLBI) [74]. This review has shown that the NAEPP guidelines were the most referenced asthma guidelines across all reviewed interventions as they were implemented in the design of 15 out of the 85 reviewed interventions [10-12,16-18,20-22,37-42].

This review identified a number of national clinical guidelines that were adopted and applied in a smaller number of interventions conducted at the national level, such as the British Guideline on the Management of Asthma [54], the Canadian Asthma Consensus Guidelines (CACG) [63], and the Air Quality Health Index (AQHI) in Canada, as well as the standardized checklist of the Dutch Asthma Foundation [57,59].

Pediatric Asthma Guidelines

In addition to the national and international guidelines, the review also identified a number of children-specific guidelines such as the Pediatric Asthma Quality of Life Questionnaire (PAQLQ) [43,50,52,53,57,61,67,68], International Survey of Asthma and Allergies in Childhood (ISAAC) questionnaire [11,12,55,68], the Eyberg Child Behavior Inventory (ECBI) [25], the Children’s Health Survey for Asthma (CHSA) by the American Academy of Pediatrics [70], and the Child Asthma Control Test (C-ACT) [52,53,68].

Other Reviews

A total of 14 other reviews of Internet- and electronic-based asthma interventions were identified. These reviews did not evaluate Internet-based asthma interventions with respect to the evidence base around the behavioral change theoretical frameworks, applied clinical guidelines, and assessment tools. However, the identified other reviews share similar discussions around main topics such as patients' perception of Internet-based interventions, limitation of existing studies, and the effect of evolving Internet and mobile technologies on the relationship between asthma patients and their health care providers.

Six of the 14 reviews indicated that Internet-based interventions were well-perceived by asthma patients and their usage was associated with promoting positive health behaviors among asthma patient groups [6,7,75-78]. On the other hand, a number of reviews reported that numerous studies for existing interventions were conducted on a small group of subjects for a limited, and often short, period of time resulting in mixed results with respect to controlling asthma symptoms and improving quality of life for asthma patients [1,7,79-81]. Last, four reviews shared concerns pertaining to the increased diffusion of Internet and mobile technologies into the delivery of care and to its impact on the clinician-patient relationship that could have negative effect on both patients and health care providers [1,7,79,80].

Discussion

Principal Findings

In an attempt to answer the primary research question pertaining to the evidence base around the behavioral change, self-care, and self-management theoretical frameworks applied within the context of the reviewed Internet-based asthma interventions, the reviewers identified 17 out of 85 interventions (20%) where at least one model, framework, and/or construct of a behavioral change theory was applied. This implies that the majority of our reviewed interventions did not apply any documented behavioral change theory to inform their design. As such, this review found that only a limited number of behavioral change theories and models were referenced and applied across multiple studies.

In total, the reviewers were able to identify 10 behavioral change theories and models that were applied across multiple (more than one) interventions versus 13 other theories and models that were only applied within the context of a single study and/or intervention.

Compared to the applied theoretical frameworks, and in response to the secondary research question, the reviewers were able to report that the application and employment of clinical guidelines and assessment tools were more salient across the reviewed interventions. The review identified six clinical guidelines that were applied across 30 of the 85 interventions (35%) as well as a total of 21 assessment tools that were applied across 32 of the 85 interventions (38%).

The National Asthma Education and Prevention Program (NAEPP) guidelines were the most referenced asthma guidelines across all the reviewed interventions as they were implemented in the design of 15 out of the 85 reviewed interventions [10-12,16-18,20-22,37-42].

Limitations

This review has a number of limitations. First, the reviewers searched the literature in six major bibliographic databases between 2004 and October 2014 only; there may be other relative studies published in other databases. Second, the reviewers did not employ any theory or guidelines to evaluate the quality of each included and reviewed study. Thus, all reviewed studies are assumed to be of the same quality.

Conclusions

It was found that the majority of published interventions did not apply behavioral change theory, clinical guidelines, and/or assessment tools to inform their design. Further, it was found that the application of clinical guidelines and assessment tools were more salient across the reviewed interventions. A consequence, therefore, is that many Internet-based asthma interventions are designed in an ad hoc manner, without the use of any notable evidence-based theoretical frameworks, clinical guidelines, and/or assessment tools.

Conflicts of Interest

The editor/publisher of this journal (GE) is academic supervisor of the first author, but had no role in making any decisions regarding this paper, which was handled by an associate editor.

References

1. McLean S, Chandler D, Nurmatov U, Liu J, Pagliari C, Car J, et al. Telehealthcare for asthma: a Cochrane review. CMAJ 2011 Aug 9;183(11):E733-E742 [FREE Full text] [doi: 10.1503/cmaj.101146] [Medline: 21746825]
2. World Health Organization. 2013 Nov. Asthma fact sheet N°307 URL: http://www.who.int/mediacentre/factsheets/fs307/en/ [accessed 2014-07-20] [WebCite Cache ID 6RDECtgui]
3. Barnett SBL, Nurmagambetov TA. Costs of asthma in the United States: 2002-2007. J Allergy Clin Immunol 2011 Jan;127(1):145-152 [FREE Full text] [doi: 10.1016/j.jaci.2010.10.020] [Medline: 21211649]
4. Smetanin P, Stiff D, Briante C, Ahmad S, Wong L, Ler A. Life and economic impact of lung disease in Ontario: 2011 to 2041. RiskAnalytica, on behalf of the Ontario Lung Association 2011 [FREE Full text]
5. Accordini S, Corsico AG, Braggion M, Gerbase MW, Gislason D, Gulsvik A, et al. The cost of persistent asthma in Europe: an international population-based study in adults. Int Arch Allergy Immunol 2013;160(1):93-101. [doi: 10.1159/000338998] [Medline: 22948386]

6. 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]
7. 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]
8. National Asthma Education and Prevention Program. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Bethesda, MD: National Heart, Lung, and Blood Institute (US); 2007.
9. Bensley RJ, Mercer N, Brusk JJ, Underhile R, Rivas J, Anderson J, et al. The eHealth Behavior Management Model: a stage-based approach to behavior change and management. Prev Chronic Dis 2004 Oct;1(4):A14 [FREE Full text] [Medline: 15670446]
10. Joseph CLM, Havstad SL, Johnson D, Saltzgaber J, Peterson EL, Resnicow K, et al. Factors associated with nonresponse to a computer-tailored asthma management program for urban adolescents with asthma. J Asthma 2010 Aug;47(6):667-673 [FREE Full text] [doi: 10.3109/02770900903518827] [Medline: 20642376]
11. Joseph CLM, Ownby DR, Havstad SL, Saltzgaber J, Considine S, Johnson D, Research team members. Evaluation of a web-based asthma management intervention program for urban teenagers: reaching the hard to reach. J Adolesc Health 2013 Apr;52(4):419-426 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.07.009] [Medline: 23299008]
12. Joseph CLM, 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]
13. Shegog R, Bartholomew LK, Gold RS, Pierrel E, Parcel GS, Sockrider MM, et al. Asthma management simulation for children: translating theory, methods, and strategies to effect behavior change. Simul Healthc 2006;1(3):151-159. [doi: 10.1097/01.SIH.0000244456.22457.e8] [Medline: 19088584]
14. Stewart M, Letourneau N, Masuda JR, Anderson S, McGhan S. Impacts of online peer support for children with asthma and allergies: "It just helps you every time you can't breathe well". J Pediatr Nurs 2013;28(5):439-452. [doi: 10.1016/j.pedn.2013.01.003] [Medline: 23398896]
15. Wyatt TH, Hauenstein EJ. Pilot testing Okay With Asthma: an online asthma intervention for school-age children. J Sch Nurs 2008 Jun;24(3):145-150. [doi: 10.1622/1059-8405(2008)024[0145:PTOWAA]2.0.CO;2] [Medline: 18557673]
16. Wyatt TH, Li X, Huang Y, Farmer R, Reed D, Burkhart PV. Developing an interactive story for children with asthma. Nurs Clin North Am 2013 Jun;48(2):271-285. [doi: 10.1016/j.cnur.2013.01.006] [Medline: 23659813]
17. Lu M, Ownby DR, Zoratti E, Roblin D, Johnson D, Johnson CC, et al. Improving efficiency and reducing costs: Design of an adaptive, seamless, and enriched pragmatic efficacy trial of an online asthma management program. Contemp Clin Trials 2014 May;38(1):19-27. [doi: 10.1016/j.cct.2014.02.008] [Medline: 24607295]
18. Hartmann CW, Sciamanna CN, Blanch DC, Mui S, Lawless H, Manocchia M, et al. A website to improve asthma care by suggesting patient questions for physicians: qualitative analysis of user experiences. J Med Internet Res 2007;9(1):e3 [FREE Full text] [doi: 10.2196/jmir.9.1.e3] [Medline: 17478412]
19. van Gaalen JL, Beerthuizen T, van der Meer V, van Reisen P, Redelijkheid GW, Snoeck-Stroband JB, et al. Long-term outcomes of internet-based self-management support in adults with asthma: randomized controlled trial. J Med Internet Res 2013;15(9):e188 [FREE Full text] [doi: 10.2196/jmir.2640] [Medline: 24028826]
20. 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]
21. Meischke H, Lozano P, Zhou C, Garrison MM, Christakis D. Engagement in "My Child's Asthma", an interactive web-based pediatric asthma management intervention. Int J Med Inform 2011 Nov;80(11):765-774 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.08.002] [Medline: 21958551]
22. Haze KA, Lynaugh J. Building patient relationships: a smartphone application supporting communication between teenagers with asthma and the RN care coordinator. Comput Inform Nurs 2013 Jun;31(6):266-71; quiz 272. [doi: 10.1097/NXN.0b013e318295e5ba] [Medline: 23728445]
23. Lin SP, Yang HY. Exploring key factors in the choice of e-health using an asthma care mobile service model. Telemed J E Health 2009 Nov;15(9):884-890. [doi: 10.1089/tmj.2009.0047] [Medline: 19919195]
24. Jeong HE, Arriaga RI. Using an ecological framework to design mobile technologies for pediatric asthma management. In: The 11th International Conference on Human-Computer Interaction with Mobile Devices and Services.: ACM; 2009 Sep Presented at: MobileHCI '09; 2009; Bonn, Germany. [doi: 10.1145/1613858.1613880]
25. Clarke SA, Calam R, Morawska A, Sanders M. Developing web-based Triple P 'Positive Parenting Programme' for families of children with asthma. Child Care Health Dev 2014 Jul;40(4):492-497. [doi: 10.1111/cch.12073] [Medline: 23662595]
26. Griffiths M. Video games and health: Video gaming is safe for most players and can be useful in health care. BMJ 2005;331(7509):122-123. [doi: 10.1136/bmj.331.7509.122]
27. Sawyer B. From cells to cell processors: the integration of health and video games. IEEE Comput Graph Appl 2008 Dec;28(6):83-85. [doi: 10.1109/MCG.2008.114] [Medline: 19004688]

28. Deterding S, Dixon D, Nacke LE, O'Hara K, Sicart M. Gamification: Using game design elements in non-gaming contexts. In: Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems. 2011 Presented at: CHI EA'11; 2011; Vancouver, BC URL: http://gamification-research.org/chi2011/ [doi: 10.1145/482.575]
29. McCallum S. Gamification and serious games for personalized health. Stud Health Technol Inform 2012;177:85-96. [Medline: 22942036]
30. Said NS. An engaging multimedia design model. In: Proceedings of the 2004 conference on Interaction Design and Children: Building a Community. New York, NY: ACM; 2004 Presented at: IDC'04; June 1-3, 2004; Baltimore, MD p. 169. [doi: 10.1145/1017833.1017873]
31. Said N, Rahman S, Yassin SF. Revisiting an engaging experience to identify metacognitive strategies towards developing a multimedia design model. International Journal of Education and Information Technologies 2009;3(2):115-125.
32. Baum A. The Health Belief Model. In: Cambridge Handbook of Psychology, Health, and Medicine. Cambridge, UK: Cambridge University Press; 1997.
33. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [Medline: 3378902]
34. 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]
35. Kreuter MW, Skinner CS. Tailoring: what's in a name? Health Educ Res 2000 Feb;15(1):1-4 [FREE Full text] [Medline: 10788196]
36. Velicer WF, Prochaska JO, Fava JL, Norman GJ, Redding CA. Smoking cessation and stress management: applications of the Transtheoretical Model of behavior change. Homeostasis in Health and Disease 1998 Jul;38(5-6):216-233.
37. Farzanfar R, Finkelstein J, Friedman RH. Testing the usability of two automated home-based patient-management systems. J Med Syst 2004 Apr;28(2):143-153. [Medline: 15195845]
38. Joshi A, Amelung P, Arora M, Finkelstein J. Clinical impact of home automated telemanagement in asthma. AMIA Annu Symp Proc 2005:1000 [FREE Full text] [Medline: 16779287]
39. Lee HR, Yoo SK, Jung SM, Kwon NY, Hong CS. A Web-based mobile asthma management system. J Telemed Telecare 2005;11 Suppl 1:56-59. [doi: 10.1258/1357633054461598] [Medline: 16035995]
40. Malone F, Callahan CW, Chan DS, Sheets S, Person DA. Caring for children with asthma through teleconsultation: "ECHO-Pac, The Electronic Children's Hospital of the Pacific". Telemed J E Health 2004;10(2):138-146. [Medline: 15319043]
41. Meltzer EO, Kelley N, Hovell MF. Randomized, cross-over evaluation of mobile phone vs paper diary in subjects with mild to moderate persistent asthma. Open Respir Med J 2008;2:72-79 [FREE Full text] [doi: 10.2174/1874306400802010072] [Medline: 19412327]
42. Van SD, Magzamen S, Truelove S, Morrison T. Remote monitoring of inhaled bronchodilator use and weekly feedback about asthma management: an open-group, short-term pilot study of the impact on asthma control. PLoS One 2013;8(2):e55335 [FREE Full text] [doi: 10.1371/journal.pone.0055335] [Medline: 23460785]
43. Deschildre A, Béghin L, Salleron J, Iliescu C, Thumerelle C, Santos C, et al. Home telemonitoring (forced expiratory volume in 1 s) in children with severe asthma does not reduce exacerbations. Eur Respir J 2012 Feb;39(2):290-296 [FREE Full text] [doi: 10.1183/09031936.00185310] [Medline: 21852334]
44. Hashimoto S, Brinke AT, Roldaan AC, van Veen IH, Möller GM, Sont JK, et al. Internet-based tapering of oral corticosteroids in severe asthma: a pragmatic randomised controlled trial. Thorax 2011 Jun;66(6):514-520. [doi: 10.1136/thx.2010.153411] [Medline: 21474498]
45. Hung SH, Tseng HC, Tsai WH, Lin HH, Cheng JH, Chang YM. Care for Asthma via Mobile Phone (CAMP). Stud Health Technol Inform 2007;126:137-143. [Medline: 17476056]
46. 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]
47. Liu WT, Huang CD, Wang CH, Lee KY, Lin SM, Kuo HP. A mobile telephone-based interactive self-care system improves asthma control. Eur Respir J 2011 Feb;37(2):310-317 [FREE Full text] [doi: 10.1183/09031936.00000810] [Medline: 20562122]
48. 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]
49. van der Meer V, van Stel HF, Bakker MJ, Roldaan AC, Assendelft WJJ, Sterk PJ, SMASHING (Self-Management of Asthma Supported by Hospitals, ICT, Nurses and General practitioners) Study Group. Weekly self-monitoring and treatment adjustment benefit patients with partly controlled and uncontrolled asthma: an analysis of the SMASHING study. Respir Res 2010;11:74 [FREE Full text] [doi: 10.1186/1465-9921-11-74] [Medline: 20537124]
50. Willems Daniëlle CM, Joore MA, Hendriks JJE, Nieman FHM, Severens JL, Wouters EFM. The effectiveness of nurse-led telemonitoring of asthma: results of a randomized controlled trial. J Eval Clin Pract 2008 Aug;14(4):600-609. [doi: 10.1111/j.1365-2753.2007.00936.x] [Medline: 19126178]

51. Uysal MA, Mungan D, Yorgancioglu A, Yildiz F, Akgun M, Gemicioglu B, Study Group, Turkish Asthma Control Test (TACT), Turkey. Asthma control test via text messaging: could it be a tool for evaluating asthma control? J Asthma 2013 Dec;50(10):1083-1089. [doi: 10.3109/02770903.2013.832294] [Medline: 23923976]
52. Voorend-van BS, Vaessen-Verberne AA, Landstra AM, Brackel HJ, van den Berg NJ, Caudri D, et al. Monitoring childhood asthma: web-based diaries and the asthma control test. J Allergy Clin Immunol 2014 Jun;133(6):1599-605.e2. [doi: 10.1016/j.jaci.2013.10.005] [Medline: 24290276]
53. Hofstede J, de Bie J, van Wijngaarden B, Heijmans M. Knowledge, use and attitude toward eHealth among patients with chronic lung diseases. Int J Med Inform 2014 Dec;83(12):967-974. [doi: 10.1016/j.ijmedinf.2014.08.011] [Medline: 25269992]
54. Ram FS, McNaughton W. Giving Asthma Support to Patients (GASP): a novel online asthma education, monitoring, assessment and management tool. J Prim Health Care 2014;6(3):238-244. [Medline: 25194251]
55. Wanick Sarinho S, Menezes Júnior J, Gusmão C, Lyra NRS. Experience in the development of a mobile diagnosis support system for asthma: Intelimed. Journal of Allergy and Clinical Immunology 2014 Feb;133(2):AB127-AB127. [doi: 10.1016/j.jaci.2013.12.474]
56. Araújo L, Jacinto T, Moreira A, Castel-Branco MG, Delgado L, Costa-Pereira A, et al. Clinical efficacy of web-based versus standard asthma self-management. J Investig Allergol Clin Immunol 2012;22(1):28-34 [FREE Full text] [Medline: 22448451]
57. Rikkers-Mutsaerts ERVM, Winters AE, Bakker MJ, van Stel HF, van der Meer V, de Jongste JC, et al. Internet-based self-management compared with usual care in adolescents with asthma: a randomized controlled trial. Pediatr Pulmonol 2012 Dec;47(12):1170-1179. [doi: 10.1002/ppul.22575] [Medline: 22644646]
58. Ryan D, Price D, Musgrave SD, Malhotra S, Lee AJ, Ayansina D, et al. Clinical and cost effectiveness of mobile phone supported self monitoring of asthma: multicentre randomised controlled trial. BMJ 2012;344:e1756 [FREE Full text] [Medline: 22446569]
59. 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]
60. van der Meer V, van Stel HF, Detmar SB, Otten W, Sterk PJ, Sont JK. Internet-based self-management offers an opportunity to achieve better asthma control in adolescents. Chest 2007 Jul;132(1):112-119. [doi: 10.1378/chest.06-2787] [Medline: 17400674]
61. Vasbinder EC, Janssens HM, Rutten-van Mölken MPMH, van DL, de Winter BCM, de Groot RCA, e-MATIC Study Group. e-Monitoring of Asthma Therapy to Improve Compliance in children using a real-time medication monitoring system (RTMM): the e-MATIC study protocol. BMC Med Inform Decis Mak 2013;13:38 [FREE Full text] [doi: 10.1186/1472-6947-13-38] [Medline: 23514242]
62. Morrison D, Wyke S, Thomson NC, McConnachie A, Agur K, Saunderson K, et al. A Randomized trial of an Asthma Internet Self-management Intervention (RAISIN): study protocol for a randomized controlled trial. Trials 2014;15:185 [FREE Full text] [doi: 10.1186/1745-6215-15-185] [Medline: 24884722]
63. Licskai C, Sands TW, Ferrone M. Development and pilot testing of a mobile health solution for asthma self-management: asthma action plan smartphone application pilot study. Can Respir J 2013 Aug;20(4):301-306 [FREE Full text] [Medline: 23936890]
64. 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]
65. Mayers I, Ohinmaa A, Vethanayagam D, Sharpe HM, O'Hara C, Majaesic C, et al. TThe virtual asthma clinic: Managing asthma patients online. Chest 2005 Oct 01;128(4_MeetingAbstracts):244S-244S. [doi: 10.1378/chest.128.4_MeetingAbstracts.244S]
66. Lin HC, Chiang LC, Wen TN, Yeh KW, Huang JL. Development of online diary and self-management system on e-Healthcare for asthmatic children in Taiwan. Comput Methods Programs Biomed 2014 Oct;116(3):299-310. [doi: 10.1016/j.cmpb.2014.05.004] [Medline: 24947614]
67. Mulvaney SA, Ho YX, Cala CM, Chen Q, Nian H, Patterson BL, et al. Assessing adolescent asthma symptoms and adherence using mobile phones. J Med Internet Res 2013;15(7):e141 [FREE Full text] [doi: 10.2196/jmir.2413] [Medline: 23864345]
68. Zomer-Kooijker K, van Erp FC, Balemans WAF, van Ewijk BE, van der Ent Cornelis K, Expert Network for Children with Respiratory and Allergic Symptoms. The expert network and electronic portal for children with respiratory and allergic symptoms: rationale and design. BMC Pediatr 2013;13:9 [FREE Full text] [doi: 10.1186/1471-2431-13-9] [Medline: 23324209]
69. Burbank AJ, Hall-Barrow J, Denman RR, Lewis SD, Hewes M, Schellhase DE, et al. There's an app for that: A pilot and feasibility study of mobile-based asthma action plans. Journal of Allergy and Clinical Immunology 2013 Feb;131(2):AB161-AB161. [doi: 10.1016/j.jaci.2012.12.1236]
70. Arnold RJ, Stingone JA, Claudio L. Computer-assisted school-based asthma management: a pilot study. JMIR Res Protoc 2012;1(2):e15 [FREE Full text] [doi: 10.2196/resprot.1958] [Medline: 23612058]

71. Cleland J, Caldow J, Ryan D. A qualitative study of the attitudes of patients and staff to the use of mobile phone technology for recording and gathering asthma data. J Telemed Telecare 2007;13(2):85-89. [doi: 10.1258/135763307780096230] [Medline: 17359572]
72. The Global Initiative for Asthma (GINA). URL: http://www.ginasthma.org/About-Us [accessed 2014-07-20] [WebCite Cache ID 6RDDusyVR]
73. The International Study of Asthma and Allergies in Childhood. URL: http://isaac.auckland.ac.nz/ [accessed 2014-07-20] [WebCite Cache ID 6RDE4e0UA]
74. The National Heart, Lung, and Blood Institute (NHLBI). 2013. National Asthma Education and Prevention Program URL: http://www.nhlbi.nih.gov/about/naepp/ [accessed 2014-07-20] [WebCite Cache ID 6RDEA8zYR]
75. 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]
76. Jaana M, Paré G, Sicotte C. Home telemonitoring for respiratory conditions: a systematic review. Am J Manag Care 2009 May;15(5):313-320 [FREE Full text] [Medline: 19435399]
77. Van Gaalen JL, Hashimoto S, Sont JK. Telemanagement in asthma: an innovative and effective approach. Curr Opin Allergy Clin Immunol 2012 Jun;12(3):235-240. [doi: 10.1097/ACI.0b013e3283533700] [Medline: 22475997]
78. Tran N, Coffman JM, Sumino K, Cabana MD. Patient reminder systems and asthma medication adherence: a systematic review. J Asthma 2014 Jun;51(5):536-543. [doi: 10.3109/02770903.2014.888572] [Medline: 24506699]
79. 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]
80. McLean S, Chandler D, Nurmatov U, Liu J, Pagliari C, Car J, et al. Telehealthcare for asthma: a Cochrane review. CMAJ 2011 Aug 9;183(11):E733-E742 [FREE Full text] [doi: 10.1503/cmaj.101146] [Medline: 21746825]
81. Marcano Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013;11:CD010013. [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]

Abbreviations

ABC: Asthma Behavior Checklist
ACQ: Asthma Control Questionnaire
ACT: Asthma Control Test
ALQ: Asthma Life Quality Questionnaire
AQHI: Air Quality Health Index
AQLQ: Asthma Quality of Life Questionnaires
ATAQ: Asthma Therapy Assessment Questionnaire
CACG: Canadian Asthma Consensus Guidelines
C-ACT: Child Asthma Control Test
CHSA: Children’s Health Survey for Asthma
COPD: chronic obstructive pulmonary disease
ECBI: Eyberg Child Behavior Inventory
eHealth: electronic health
EHR: electronic health records
EMR: electronic medical records
GINA: Global Initiative for Asthma guidelines
HIS: hospital information system
IMS: Illness Management Survey
ISAAC: International Survey of Asthma and Allergies in Childhood questionnaire
KASE-AQ: Knowledge, Attitude and Self-Efficacy Asthma Questionnaire
Mini AQLQ: Mini Asthma Quality of Life
NAEPP: National Asthma Education and Prevention Program
NHLBI EP3: The National Heart, Lung, and Blood Institute, Expert Panel Report 3
NHLBI: National Heart, Lung, and Blood Institute
PAQLQ: Pediatric Asthma Quality of Life Questionnaire
SMS: short message service
UX: user experience
WDTA: Watch-Discover-Think-Act
WHO: World Health Organization

Edited by H Potts; submitted 08.12.14; peer-reviewed by J Fonseca, K Stawarz; comments to author 31.12.14; revised version received 09.02.15; accepted 09.03.15; published 02.04.15.

Please cite as:
Al-Durra M, Torio MB, Cafazzo JA
The Use of Behavior Change Theory in Internet-Based Asthma Self-Management Interventions: A Systematic Review
J Med Internet Res 2015;17(4):e89
URL: http://www.jmir.org/2015/4/e89/
doi:10.2196/jmir.4110
PMID:25835564

Original Paper

Long-Term Outcomes of a Web-Based Diabetes Prevention Program: 2-Year Results of a Single-Arm Longitudinal Study

S Cameron Sepah[1,2], PhD; Luohua Jiang[3], PhD; Anne L Peters[4], MD

[1]University of California, San Francisco, Department of Psychiatry, San Francisco, CA, United States
[2]Omada Health, San Francisco, CA, United States
[3]University of California, Irvine, Department of Epidemiology, Irvine, CA, United States
[4]University of Southern California, Keck School of Medicine, Los Angeles, CA, United States

Corresponding Author:
S Cameron Sepah, PhD
University of California, San Francisco
Department of Psychiatry
401 Parnassus Ave, Box 0984
San Francisco, CA, 94143
United States
Phone: 1 888 987 8337
Fax: 1 855 284 5113
Email: cameron.sepah@ucsf.edu

Abstract

Background: Digital therapeutics are evidence-based behavioral treatments delivered online that can increase accessibility and effectiveness of health care. However, few studies have examined long-term clinical outcomes of digital therapeutics.

Objective: The objective of this study was to conduct a 2-year follow-up on participants in the Internet-based *Prevent* diabetes prevention program pilot study, specifically examining the effects on body weight and A1c, which are risk factors for diabetes development.

Methods: A quasi-experimental research design was used, including a single-arm pre- and post-intervention assessment of outcomes. Participants underwent a 16-week weight loss intervention and an ongoing weight maintenance intervention. As part of the program, participants received a wireless scale, which was used to collect body weight data on an ongoing basis. Participants also received A1c test kits at baseline, 0.5 year, 1 year, and 2-year time points.

Results: Participants previously diagnosed with prediabetes (n=220) were originally enrolled in the pilot study. A subset of participants (n=187) met Centers for Disease Control and Prevention (CDC) criteria for starting the program (starters), and a further subset (n=155) met CDC criteria for completing the program (completers) and were both included in analyses. Program starters lost an average of 4.7% (SD 0.4) of baseline body weight after 1 year and 4.2% (SD 0.8) after 2 years, and reduced A1c by mean 0.38% (SD 0.07) after 1 year and 0.43% (SD 0.08) after 2 years. Program completers lost mean 4.9% (SD 0.5) of baseline body weight after 1 year and 4.3% (SD 0.8) after 2 years, and reduced A1c by 0.40% (SD 0.07) after 1 year and 0.46% (SD 0.08) after 2 years. For both groups, neither 2-year weight loss nor A1c results were significantly different from 1-year results.

Conclusions: Users of the *Prevent* program experienced significant reductions in body weight and A1c that are maintained after 2 years. Contrary to the expected progression from prediabetes to diabetes over time, average A1c levels continued to show an average regression from within the prediabetic range (5.7%-6.4%) initially to the normal range (<5.7%) after 2 years. Further investigation is warranted to test digital therapeutics as a scalable solution to address national diabetes and cardiovascular disease prevention efforts.

(J Med Internet Res 2015;17(4):e92) doi:10.2196/jmir.4052

KEYWORDS

prediabetes; type 2 diabetes; obesity; diabetes prevention program; internet; online; mobile apps; mhealth; digital health; intervention

Introduction

Prediabetes, the clinical precursor to type 2 diabetes, continues to grow to epidemic levels. Recent estimates by the Centers for Disease Control and Prevention (CDC) indicate that prevalence increased 8% in the last decade—from 29% in 1999-2002 to 37% in 2009-2012—amounting to 86 million Americans over age 20 with prediabetes [1,2]. Furthermore, the US Preventive Services Task Force (USPSTF) now recommends that overweight/obese adults with additional cardiovascular risk factors such as prediabetes receive intensive behavioral counseling interventions such as the Diabetes Prevention Program (DPP) [3]. As a result, there is an urgent need for effective chronic disease prevention interventions that can be disseminated to millions of people. To address such needs, a new field of "digital therapeutics" has emerged—evidence-based treatments from the field of behavioral medicine that are delivered online [4].

One of the most established digital therapeutics is *Prevent*, an Internet-based version of the DPP clinical trial. Though the DPP was primarily an individual treatment, most DPP translations including *Prevent* have successfully used group-based approaches to minimize cost, and thus group-based interventions have been recommended in evidence-based guidelines for diabetes prevention [5].

Previous published research showed that *Prevent* program starters were able to achieve clinically significant reductions in body weight and A1c after 1 year [6]. Of note, this was the first study to show that a scalable Internet-based intervention met all efficacy benchmarks for DPP programs set by CDC's Diabetes Prevention and Recognition Program (DPRP) standards [7]. This addressed a key concern raised by Yudkin & Montori that "rolling out intensive lifestyle interventions like these to populations with pre-diabetes…would be challenging" [8].

Furthermore, a critique of DPP studies by Kahn and Davidson noted that weight loss results of DPP translations have often been low (a meta-analysis of 22 studies showing 2.4% weight loss at 1 year) [9] and often show significant weight regain after the intervention ends. However, they concede that real-world DPPs "may have promise if the weight loss achieved at year 2 can be sustained and replicated" [10].

Thus, the current study seeks to make an original contribution to the scientific literature by addressing such points through investigating the long-term outcomes and sustainability of an Internet-based DPP.

Methods

Research Design

A quasi-experimental research design was used, including a single-arm pre- and post-intervention assessment of body weight, A1c, and program engagement outcomes [6]. Participants were not compensated for participation but were enrolled in the program at no cost. Institutional Review Board (IRB) exemption was granted by Western IRB for secondary analyses of previously collected and de-identified data.

Participants

Patients were recruited via craigslist advertisements seeking participants for an Internet-based diabetes prevention program. Patients were screened for a self-reported clinical diagnosis of prediabetes within the past year and meeting CDC DPRP eligibility criteria: 18 years of age or older, have a body mass index (BMI) of ≥24 kg/m^2 (≥22 kg/m^2 if Asian), and able to engage in light physical activity [7].

Eligible participants completed an online account set-up process, in which they provided consent and completed health and demographic questionnaires, and then enrolled in the *Prevent* program, which they could access via any Internet-enabled desktop or mobile device [11].

Program

A full description of the program components has been previously published [6]. Briefly, *Prevent* is an Internet-based translation of the DPP lifestyle intervention, which includes small group support, personalized health coaching, a weekly DPP-based curriculum, and digital tracking tools. Participants were demographically matched into groups of 10-15 participants and placed into a private online social network resembling Facebook where they could discuss goal progress and provide social support to one another. At any convenient time or place using Internet-enabled devices, they could asynchronously complete weekly DPP-based health education lessons, privately message and call a health coach for individual counseling, track weight loss, and physical activity using a wireless weight scale and pedometer, and monitor their engagement and weight loss progress.

Prevent starts with a 16-week "Core" curriculum focusing on weight loss and continues with a 36-week post-core "Sustain" curriculum focusing on weight maintenance, with this "active" intervention totaling 12 months. During this time, participants also engage with health coaching, small group discussion, and tracking of body weight/food/physical activity. After 12 months, participants continue with an ongoing intervention that is more proactive, in which they continue to have access to past curriculum, a larger *Prevent* "alumni" group discussion forum, and tracking capabilities.

Measures

Baseline demographic and health information were collected prior to program start using an Internet-based questionnaire self-completed by the subject. Body weight, the study's primary outcome, was measured serially in pounds using a validated, wireless-enabled scale that was mailed to participants [12]. Weights were transmitted over a 2G cellular network to a central study server. Participants were encouraged to weigh themselves daily and reminded via email/telephone calls to obtain weights at the baseline, 6, 12, and 24-month assessment timepoints. Weight measurements were highly stable and the scale's coefficient of variation was ±0.2 lb.

A1c was measured in percent (NGSP/DCCT units) using self-administered AccuBase A1c test kits by DTI laboratories, an FDA-cleared whole blood test that uses a capillary tube blood collection method. This allows for reliable home-based data collection and valid lab testing using high-performance liquid chromatography (HPLC-IE/HPLC-BA), including abnormal hemoglobin screening [13]. A1c test kits were mailed to participants' homes around the baseline, 6, 12, and 24-month assessment timepoints.

Analyses

Results were analyzed for two subgroups based on CDC DPRP standards: "program starters" were those who completed at least 4 core lessons, and "program completers" were those who completed as least 9 core lessons.

Results were analyzed using SPSS Statistics 21.0 and SAS 9.3. Baseline characteristics were compared between subgroups using chi-square tests or Fisher's exact test for categorical variables and two-sample *t* tests for continuous variables. To account for repeated measures and missing data, linear mixed-effects models were used to obtain adjusted mean changes in weight and A1c over the 2-year follow-up period. Based on exploratory data analysis and graphs of the time trends, piecewise linear models were fit for weight change with days from baseline and a change point after the last core lesson day included in the model. For A1c, an additional change point at 12 months was added to the mixed model due to significant change in slope at that timepoint and better model fit statistics. Models used an autoregressive-moving-average covariance structure to statistically account for the correlation of frequently measured weight data. Repeated A1c measures were also correlated but not measured on a daily basis. Therefore a spatial power covariance structure (with time as the distance measure) was used to account for the correlation among repeated measures of A1c from the same participant.

Results

Demographics and Participation

As shown in Figure 1, 254 participants from across the United States responded to online advertisements and met CDC DPRP eligibility criteria. Of these, 220 participants completed the initial assessment and online set-up process and enrolled in the intervention on April 29, 2012. Demographic characteristics of study participants are reported in a previous study [5]. Briefly, as outlined in Table 1, the participants were socioeconomically diverse. The baseline BMI of these 220 participants was 36.6 kg/m^2. A subset of these participants (n=187) completed at least 4 core lessons and were thus designated program starters, and a further subset of these participants (n=155) completed at least 9 core lessons and were thus designated program completers.

Figure 1. Participant recruitment and retention flow chart.

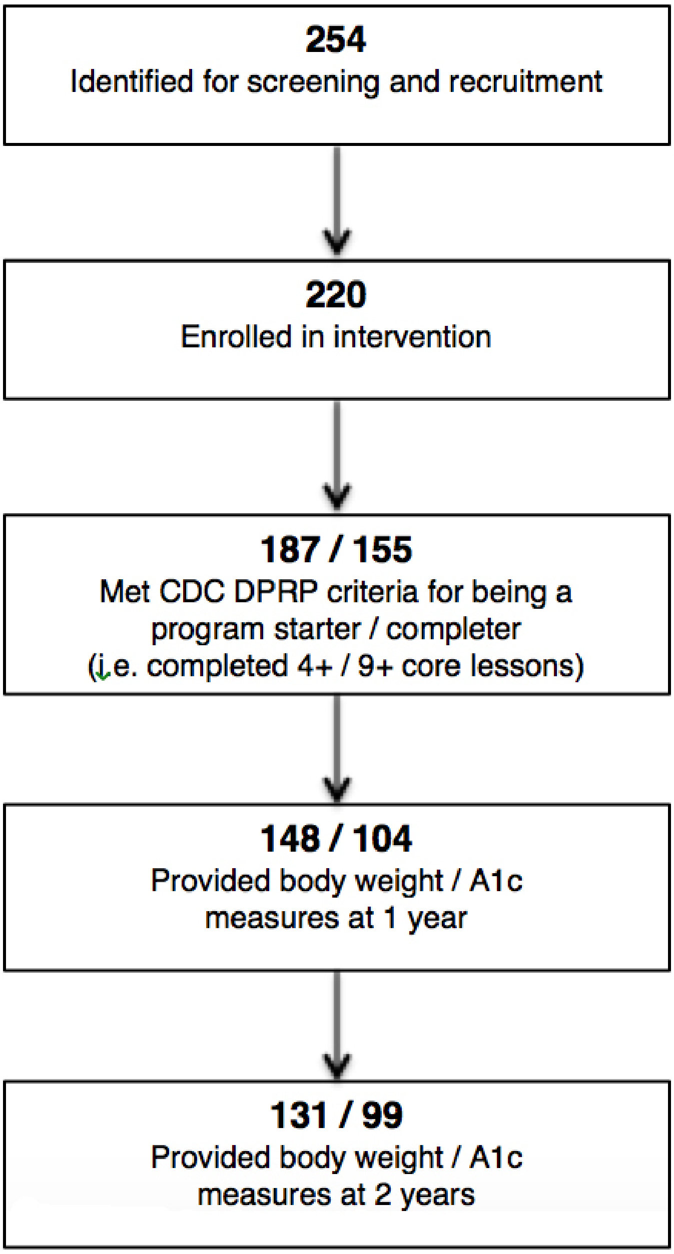


Table 1. Baseline demographic characteristics of participants.

Characteristics	Total (N=220)	Non-starters (0-3 lessons) (n=33)	Starters (4+ lessons) (n=187)	Non-starters vs Starters, *P* value	Non-completers (4-8 lessons) (n=32)	Completers (9+ lessons) (n=155)	Non-completers vs Completers, *P* value
Age, mean (SD)	43.6 (12.4)	42.0 (12.6)	43.9 (12.4)	.43[a]	39.0 (9.4)	44.9 (12.8)	.004[a]
Weight, mean (SD)	223.1 (47.9)	226.1 (53.5)	222.5 (47.0)	.69[a]	229.8 (45.9)	221.0 (47.2)	.34[a]
BMI, mean (SD)	36.6 (7.5)	35.9 (6.6)	36.7 (7.6)	.56[a]	38.3 (7.5)	36.4 (7.6)	.21[a]
Gender (male), n (%)	38 (17.3)	10 (30.3)	28 (15.0)	.03[b]	3 (9.4)	25 (16.1)	.42[c]
Ethnicity, n (%)				.28[b]			.89[c]
White	108 (50.2)	15 (45.5)	93 (51.1)		15 (48.4)	78 (51.7)	
Black	63 (29.3)	10 (30.3	53 (29.1)		11 (35.5)	42 (27.8)	
Hispanic	23 (10.7)	2 (6.1)	21 (11.5)		3 (9.7)	18 (11.9)	
Other	21 (9.8)	6 (18.2)	15 (8.2)		2 (6.5)	13 (8.6)	
Marital status, n (%)				.13[c]			.68[c]
Married/live with a partner	87 (57.6)	9 (50.0)	78 (58.6)		15 (68.2)	63 (56.8)	
Divorced/separated/widowed	25 (16.6)	1 (5.6)	24 (18.1)		3 (13.6)	21 (18.9)	
Never married	39 (25.8)	8 (44.4)	31 (23.3)		4 (18.2)	27 (24.3)	
Education, n (%)				.01[b]			.13[b]
< College graduate	72 (48.3)	14 (77.8)	58 (44.3)		13(59.1)	45 (41.3)	
≥ College graduate	77 (51.7)	4 (22.2)	73 (55.7)		9 (40.9)	64 (58.7)	
Income (USD), n (%)				.92[b]			.52[b]
<$50,000	69 (48.3)	8 (47.1)	61 (48.4)		11 (55.0)	50 (47.2)	
$50,000 or higher	74 (51.8)	9 (52.9)	65 (51.6)		9 (45.0)	56 (52.8)	

[a] *P* value of 2-sample *t* test.

[b] *P* value of chi-square test unless otherwise noted.

[c] *P* value of Fisher's exact test.

Engagement

As reported in a previous publication, during the first year of the program, program starters completed an average of 13.8 and 3.2 lessons during the core and post-core phases (CDC DPRP benchmark is 9 and 3 lessons), documented body weight at 90% and 67% of weeks and months in which core and post-core lessons were completed (benchmark is 80% and 60%), and documented physical activity at 85% of weeks in which core lessons were completed (benchmark is 80%) [6].

Because lessons are limited to the first year, no engagement benchmarks exist beyond 1 year. However, participants continued to weigh in an average of 6.2 (SD 0.3) of 12 months and logged in an average of 3.5 (SD 0.3) of 12 months between years 1 and 2. In order to evenly compare engagement rates between years 1 and 2, the last 8 months of the year 1 (ie, the Sustain post-core period) was compared to the first 8 months of year 2. *Prevent* program starters weighed in an average of 4.6 (SD 0.2) of the last 8 months of year 1 versus 4.5 (SD 0.2) of the first 8 months of year 2, a non-significant difference (*P*=.708). *Prevent* program starters logged in on an average of 3.5 of 8 (SD 0.3) last months of the year 1 versus 2.7 of 8 (SD 0.3) first months of year 2, which was a significant difference (*P*=.008).

Changes in Body Weight and A1c

Among program starters, 100% (187/187) had an initial baseline weight measurement, 78.6% (147/187) had at least one weight measurement between 15 and 17 weeks, 79.1% (148/187) had at least one weight measurement between 11 and 13 months, and 70.1% (131/187) had at least one weight measurement between 22 and 26 months.

Changes in body weight and A1c after 16 weeks and 1 year were reported in a previous publication and reported here alongside 2-year data in Tables 2 and 3 [6]. From baseline to year 2, program starters and completers achieved mean weight loss of 4.2% (SD 0.8%) and 4.3% (SD 0.8%), respectively. Program starters' and completers' weight loss was also maintained from year 1 to year 2, with no significant change (*P*=.25 and .20, respectively). Intention-to-treat analyses were also conducted on all enrollees: from baseline to year 2, enrollees achieved an average weight loss of 4.5% (SD 0.7%), though the majority of these additional participants did not weigh in after 1 year, which contributed to a higher possibility of model error. Among the program starters who reported weight between 22 and 26 months (n=131), 52 (40.0%) continued to meet or exceeded the CDC DPRP 5% weight loss benchmark at 24 months.

Because A1c measurement was optional, compliance was lower. Among program starters, 75.4% (141/187) had an initial baseline A1c measurement, 36.4% (68/187) had an A1c measurement between months 6-8 (due to A1c being a lagging measurement), 55.6% (104/187) had an A1c measurement between months 12-14, and 52.9% (99/187) had an A1c measurement between months 24-28.

As shown in Tables 2 and 3, from baseline to year 2, program starters and completers reduced A1c by 0.43% (SD 0.08) and 0.46% (SD 0.08), respectively. Program starters' and completers' A1c reduction was also maintained from year 1 to year 2, with no significant change (*P*=.39 and .38, respectively).

Table 2. Body weight and A1c of participants over time.

	Starters (4+ lessons)		Completers (9+ lessons)	
	Weight (lbs), mean (SE)[a]	A1c (%), mean (SE)[a]	Weight (lbs), mean (SE)[a]	A1c (%), mean (SE)[a]
Baseline	221.6 (3.5)	5.99 (0.07)	220.2 (3.9)	6.02 (0.08)
16 weeks	210.5 (3.5)	6.02 (0.08)	208.6 (3.9)	6.04 (0.09)
Year 1	211.3 (3.4)	5.61 (0.08)	209.5 (3.8)	5.62 (0.08)
Year 2	212.3 (3.5)	5.55 (0.08)	210.7 (3.9)	5.56 (0.08)

[a] Adjusted means from linear mixed models.

Table 3. Changes in body weight and A1c of participants over time.

	Starters (4+ lessons)				Completers (9+ lessons)			
	Weight loss, % change (SE)[a]	*P* value	A1c, change (SE)[a]	*P* value	Weight loss, % change (SE)[a]	*P* value	A1c, change (SE)[a]	*P* value
16 weeks-Baseline	5.0 (0.3)	<.001	0.03 (0.06)	.55	5.2 (0.3)	<.001	0.03 (0.06)	.62
Year 1-Baseline	4.7 (0.4)	<.001	-0.38 (0.07)	<.001	4.9 (0.5)	<.001	-0.40 (0.07)	<.001
Year 2-Baseline	4.2 (0.8)	<.001	-0.43 (0.08)	<.001	4.3 (0.8)	<.001	-0.46 (0.08)	<.001
Year 2-Year 1	-0.5 (-0.4)	.25	-0.06 (0.07)	.39	-0.5 (-0.5)	.20	-0.06 (0.07)	.38

[a] Adjusted means from linear mixed models.

Discussion

Principal Findings

Results indicate that the *Prevent* Internet-based diabetes prevention program helped participants who started or completed the program achieve significant reductions in body weight and A1c after 2 years. Of note, weight loss was largely maintained between years 1 and 2; this is one of the few studies to exhibit this weight maintenance effect after the active intervention ended.

Furthermore, A1c continued to show an average reduction from within the prediabetes range (5.7%-6.4%) to the normal range (<5.7%), in contrast to an expected annual rate of progression of approximately 4% from prediabetes to type 2 diabetes [14]. Thus, these results address critiques regarding the ability of DPP translations to show long-term effects on weight loss and A1c reduction [7,9].

The ability of *Prevent* to show sustained effects may be due to unique aspects of the program. Although the active intervention was 12 months, participants have continued access to *Prevent* and can proactively use the intervention. Thus, by allowing ongoing health tracking and social support, digital therapeutics can promote long-term maintenance of lifestyle changes with little ongoing intervention cost or maintenance.

Although *Prevent* exceeded CDC DPRP benchmarks for program completion during the 16-week core phase, 30% of participants who initially enrolled still did not complete the program. We hypothesize that one of the key advantages of digital therapeutics—accessibility—may also be a disadvantage when it comes to program completion. The lower barrier to Internet-based program entry inevitably increases the number of participants with lower motivation, who may have never made the effort to show up to an in-person program.

Strengths and Limitations

Study strengths include longitudinal collection of body weight and A1c data and statistical analysis using linear mixed models, which allow for more robust estimation over time and missing data. Furthermore, recruitment, intervention, and assessment were done exclusively remotely, in contrast to most digital health studies that require in-person orientation or follow-up assessment. This enhances the generalizability of the findings to "real-world" commercial deployments that must be done remotely.

Study limitations include a non-randomized, uncontrolled single-arm design with a self-selected sample, which precludes causal inference of the intervention to outcomes. However, this also better approximates how commercial programs enroll real-world populations. Fewer males participated in the study, but this is typical of behavioral weight loss interventions. Furthermore, 70% of program starters had weight data at year 2 and 53% had A1c data at year 2, which limits generalizability regarding outcomes on all participants. Thus, conclusions regarding weight loss and A1c reductions are limited to program starters and completers. In addition, while adherence to program behaviors (eg, lesson completion, tracking of weight and physical activity) were assessed according to CDC DPRP standards, adherence to health behaviors (eg, diet and exercise goals) were not, which limits causal inference.

Conclusions

Results of this study suggest that the *Prevent* Internet-based diabetes prevention program was able to produce significant reductions in body weight and A1c that are sustained over the course of 2 years, even after active intervention ended after 1 year. Further investigation is warranted to test digital therapeutics as a scalable solution to address national diabetes and cardiovascular disease prevention efforts.

Acknowledgments

The authors wish to acknowledge Dr Ann Albright and Dr Ed Gregg at the CDC for their support of this work, and the Diabetes Prevention Program Research Group for performing the original DPP trial, on which the *Prevent* curriculum is based.

Authors' Contributions

Data were collected by SCS and provided unmodified to LJ, who independently conducted data analysis and wrote the results section. ALP provided consultation and editing of the paper for accuracy.

Conflicts of Interest

This work was funded by Omada Health, a company that makes and owns online behavior change programs, including the *Prevent* program, which is subject of this study. SCS is medical director for Omada Health and receives a salary and stock options. LJ is a paid statistical consultant, and ALP is a scientific advisor for Omada Health and receives stock options.

References

1. Bullard KM, Saydah SH, Imperatore G, Cowie CC, Gregg EW, Geiss LS, et al. Secular changes in U.S. Prediabetes prevalence defined by hemoglobin A1c and fasting plasma glucose: National Health and Nutrition Examination Surveys, 1999-2010. Diabetes Care 2013;36(8):2286-2293 [FREE Full text] [doi: 10.2337/dc12-2563] [Medline: 23603918]
2. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2014 URL: http://www.cdc.gov/diabetes/pubs/statsreport14/national-diabetes-report-web.pdf [accessed 2014-11-17] [WebCite Cache ID 6UAMyRwZ4]
3. LeFevre ML, US Preventive Services Task Force. Behavioral counseling to promote a healthful diet and physical activity for cardiovascular disease prevention in adults with cardiovascular risk factors: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med 2014;161(8):587-593. [doi: 10.7326/M14-1796] [Medline: 25155419]
4. 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]
5. Paulweber B, Valensi P, Lindström J, Lalic NM, Greaves CJ, McKee M, et al. A European evidence-based guideline for the prevention of type 2 diabetes. Horm Metab Res 2010;42 Suppl 1:S3-36 [FREE Full text] [doi: 10.1055/s-0029-1240928] [Medline: 20391306]
6. Sepah SC, Jiang L, Peters AL. Translating the Diabetes Prevention Program into an Online Social Network: Validation against CDC Standards. Diabetes Educ 2014;40(4):435-443 [FREE Full text] [doi: 10.1177/0145721714531339] [Medline: 24723130]
7. Centers for Disease Control and Prevention. 2011. Diabetes Prevention Recognition Program Standards and Operating Procedures URL: http://www.cdc.gov/diabetes/prevention/pdf/DPRP_Standards_09-02-2011.pdf [accessed 2014-11-18] [WebCite Cache ID 6UANEXBIb]
8. Yudkin JS, Montori VM. The epidemic of pre-diabetes: the medicine and the politics. BMJ 2014;349:g4485. [doi: 10.1136/bmj.g4485] [Medline: 25028385]
9. Dunkley AJ, Bodicoat DH, Greaves CJ, Russell C, Yates T, Davies MJ, et al. Diabetes prevention in the real world: effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes and of the impact of adherence to guideline recommendations: a systematic review and meta-analysis. Diabetes Care 2014 Apr;37(4):922-933. [doi: 10.2337/dc13-2195] [Medline: 24652723]
10. Kahn R, Davidson MB. The reality of type 2 diabetes prevention. Diabetes Care 2014;37(4):943-949. [doi: 10.2337/dc13-1954] [Medline: 24652724]

11. Prevent by Omada Health. URL: http://preventnow.com/ [accessed 2014-12-21] [WebCite Cache ID 6UyjAqoDN]
12. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. Daily self-weighing and adverse psychological outcomes: a randomized controlled trial. Am J Prev Med 2014;46(1):24-29. [doi: 10.1016/j.amepre.2013.08.006] [Medline: 24355668]
13. DTI Laboratories, Inc. Technical Data. URL: http://dtilaboratories.com/new_page_9.htm [accessed 2015-01-02] [WebCite Cache ID 6VH4AXs5s]
14. Morris DH, Khunti K, Achana F, Srinivasan B, Gray LJ, Davies MJ, et al. Progression rates from HbA1c 6.0-6.4% and other prediabetes definitions to type 2 diabetes: a meta-analysis. Diabetologia 2013 Jul;56(7):1489-1493. [doi: 10.1007/s00125-013-2902-4] [Medline: 23584433]

Abbreviations

BMI: Body Mass Index
CDC: Centers for Disease Control and Prevention
DPP: Diabetes Prevention Program
DPRP: Diabetes Prevention Recognition Program
HPLC: High-Performance Liquid Chromatography
IRB: Institutional Review Board
USPSTF: US Preventive Services Task Force

Edited by G Eysenbach; submitted 17.11.14; peer-reviewed by L Saslow, K Schneider; comments to author 11.12.14; revised version received 02.01.15; accepted 16.03.15; published 10.04.15.

Please cite as:
Sepah SC, Jiang L, Peters AL
Long-Term Outcomes of a Web-Based Diabetes Prevention Program: 2-Year Results of a Single-Arm Longitudinal Study
J Med Internet Res 2015;17(4):e92
URL: http://www.jmir.org/2015/4/e92/
doi:10.2196/jmir.4052
PMID:25863515

Original Paper

A Mixed-Methods Study on the Acceptability of Using eHealth for HIV Prevention and Sexual Health Care Among Men Who Have Sex With Men in China

Kathryn E Muessig[1], AB, PhD; Cedric H Bien[2], BS; Chongyi Wei[3], MA, DrPH; Elaine J Lo[1], MPH; Min Yang[4]; Joseph D Tucker[5], MA, MD, PhD; Ligang Yang[6], MD; Gang Meng[7], BS; Lisa B Hightow-Weidman[5], MPH, MD

[1]Department of Health Behavior, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Icahan School of Medicine at Mount Sinai, New York, NY, United States

[3]Department of Epidemiology and Biostatistics, School of Medicine, University of California, San Francisco, San Francisco, CA, United States

[4]Department of Health Policy and Management, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[5]Division of Infectious Diseases, School of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[6]Guangdong Provincial Center for Skin Diseases and STI Control & Prevention, Guangzhou, China

[7]Guangtong – Lingnan MSM Community Support Center, Guangzhou, China

Corresponding Author:
Kathryn E Muessig, AB, PhD
Department of Health Behavior
Gillings School of Global Public Health
The University of North Carolina at Chapel Hill
306 Rosenau Hall
Campus Box 7440
Chapel Hill, NC, 27599
United States
Phone: 1 919 962 5059
Fax: 1 919 966 2921
Email: kmuessig@med.unc.edu

Abstract

Background: Human immunodeficiency virus (HIV) infection disproportionately affects men who have sex with men (MSM). Over half of all HIV-positive MSM in China may not know their HIV status. Mobile phones and Web interventions (eHealth) are underutilized resources that show promise for supporting HIV education, testing, and linkage to care.

Objective: This mixed-methods study among MSM in China assessed technology utilization and eHealth acceptability for sexual health care.

Methods: We conducted in-depth interviews and an online survey. Qualitative analyses informed the development of the Internet survey, which was administered through two popular MSM websites. Bivariate and multivariate analysis assessed characteristics of MSM interested in eHealth for sexual health care.

Results: The qualitative sample included MSM across a range of ages, education, marital status, sexuality, and HIV testing experience. Qualitative findings included the importance of the Internet as the primary source of information about sexual health, HIV and other sexually transmitted diseases (STDs), use of the Internet to enable HIV testing opportunities by facilitating connections with both the gay community and health care providers, and mixed perceptions regarding the confidentiality of eHealth tools for sexual health. Among the Internet sample (N=1342), the average age was 30.6 years old, 82.81% (1098/1342) were single, and 53.42% (711/1331) had completed college. In the past 3 months, 38.66% (382/988) had condomless sex and 60.53% (805/1330) self-reported having ever tested for HIV. The majority of men owned computers (94.14%, 1220/1296) and mobile phones (92.32%, 1239/1342), which many had used to search for HIV/STD information and testing sites. In multivariate analysis, interest in using computers or mobile phones to support their sexual health care was associated with being a student, prior use of computers or mobile phones to search for general health information, prior use of computers or mobile phones to search for HIV/STD information, and confidentiality concerns.

Conclusions: MSM in this sample had high utilization of technology and interest in eHealth despite confidentiality concerns. Future eHealth interventions can thoughtfully and creatively address these concerns as a priority for successful implementation.

(J Med Internet Res 2015;17(4):e100) doi:10.2196/jmir.3370

KEYWORDS

Internet; HIV; AIDS; China; men who have sex with men; mixed method; sexually transmitted diseases

Introduction

Many men with human immunodeficiency virus (HIV) and who have sex with men (MSM) do not know their HIV status—ranging from around 21% in the United States [1] and New Zealand [2], to over 90% in Malawi [3]. In 2012, China's national HIV prevalence among MSM reached 6.7% but exceeded 15% in some urban areas [4]. A high national syphilis prevalence around 12% among MSM [5], combined with entrenched fear, stigma, and misunderstanding surrounding HIV and same-sex behavior, inhibit intervention efforts and fuel continued transmission. While government efforts have increased voluntary counseling and testing [6], in 2011 only half of MSM had received an HIV test in the past year and knew their results and as many as 86% of HIV-positive MSM were unaware of their infection [7]. After receiving a positive diagnosis, MSM face delays in care as they pursue required confirmatory and CD4 cell count testing, which are often available only in locations different from the place of original testing and eventual treatment. In this decentralized system, less than half of MSM receive a CD4 count within 6 months of diagnosis [8,9] and under a quarter who qualify for antiretroviral therapy (ART) initiate timely treatment [9]. Delayed linkage to care and suboptimal retention in care hinder effective ART, threaten individual patient outcomes, and may increase the likelihood of onward HIV transmission [10-16].

The tools provided by widespread Internet access and mobile phone ownership offer a promising approach to reducing these gaps in HIV prevention, testing, and care. eHealth approaches support health care services via a wide range of information and communication technologies including software and hardware, Internet and mobile phone-based resources, and global positioning system (GPS) and tablet technologies. Internet and mobile phone-based components have been successfully used within prevention and care initiatives for HIV and other sexually transmitted diseases (STD) [17-19] including promoting prevention messages [20], facilitating test result notification [21,22], and improving ART adherence and attendance at clinic appointments [23-28]. To date, phone-based HIV interventions have primarily utilized voice or text messaging (SMS) functions [19,25,29,30], but smartphones may further facilitate the delivery of more complex, interactive, and tailored interventions via the mobile Web [31,32] and software apps [33,34]. Smartphones facilitate affordable, widespread access to the Internet, narrowing the "digital divide" such that, by 2011, 68% of Chinese mobile phone owners had smartphones [35].

MSM-tailored eHealth initiatives are gaining traction in the United States [31,32,34] but have yet to be systematically designed and tested internationally. The hidden nature of MSM in Chinese mainstream society alongside widespread ownership of mobile technology and the growth of sex seeking websites and online dating services [36,37] make mobile technologies an appropriate, familiar medium to address sexual risk and health promotion. Preliminary survey research suggests that MSM in China are willing to provide Internet and mobile phone personal contact information [38]. In order to develop eHealth interventions specifically for Chinese MSM, additional information is needed about Internet and mobile phone ownership, usage, and preferences for sexual health-related services. Building on eHealth lessons learned in development research with at-risk MSM in the United States [31,32], this mixed-methods study explored how Chinese MSM are currently using the Web and mobile Web to find health information and specifically information related to sexual health. We also aimed to characterize men who are currently willing to use eHealth for sexual health care and to identify possible concerns and barriers to using eHealth. The ultimate goal for these research findings is to inform the design of eHealth interventions to address HIV prevention and sexual health care among MSM in China.

Methods

Part I: Qualitative Research

From May to August 2012, we conducted qualitative interviews in two large cities in South China—an area that bears a disproportionate burden of HIV, syphilis, and other STDs especially among MSM [39-43]. The data presented in this manuscript are part of a larger project on MSM testing preferences and experiences. We recruited individuals 16 years of age and older who reported being born biologically male and ever having sex with another man. Participants were recruited through government-sponsored health clinics, non-governmental and community-based organizations, and online through a popular MSM website. We purposively [44] enrolled men across a range of ages, education, marital status, sexuality, and HIV/STD testing experience (never tested, first time being tested, tested multiple times).

Qualitative interviews included questions on sexual health care and HIV/syphilis testing experiences and preferences. Technology questions addressed mobile phone ownership and capabilities, Internet and mobile phone use, experiences and preferences for using technology for general- and HIV/STD-related health care. The interview guide was pre-tested in mock interviews with volunteers at MSM community-based testing centers and subsequently refined. Eligible participants were interviewed one-on-one by trained, bilingual study staff in private rooms at participating study sites. Interviews were conducted in Mandarin, Cantonese, or English, and recorded with permission. Mobile phone cards and supermarket gift cards worth approximately US $10 were offered as remuneration.

All qualitative interviews were transcribed verbatim. The analysis team (KEM, CHB, EJL, JDT) developed a structured codebook to help define and illustrate a priori and emergent themes. Two study team members (CHB, EJL) were trained in how to use the codebook and then independently coded all interview transcripts using Atlas.ti, version 7.0. These coded transcripts were merged and reviewed by a third coder (KEM) for fidelity to the codebook and consistency between coders. No major differences in code applications were identified. All relevant coded data were then reviewed theme by theme and relationships between themes were described. Illustrative quotes were chosen by group consensus to present typical responses and variation within each theme. The findings are organized into four sections as follows: qualitative sample characteristics, the Internet as a primary source of information about sexual health and HIV/STDs, the role of the Internet in facilitating HIV/STD testing, and concerns and considerations for using eHealth for sexual health care among MSM in China.

Part II: Internet Survey

The qualitative findings informed the design of an Internet-based survey. Survey development was supported by sociologists who conduct the Chinese national survey of sexual behavior [45,46]. The questions were reviewed by four local MSM community volunteers and one focus group discussion. The online survey administration system was pilot tested among 201 MSM (data not included in final analysis). The final survey tool contained 225 possible questions and took an average of 20 minutes to complete. In addition to sociodemographic information, the survey content covered six broad topics: (1) condom and lubrication preferences (results not shown), (2) drug and alcohol use (results not shown), (3) sexual behavior, (4) HIV/STD testing, (5) technology ownership/use, and (6) interest in using eHealth tools for sexual health and confidentiality concerns of using eHealth for sexual health.

In May 2013, the Internet survey was hosted by two high-traffic Web portals: GZTZ.org (Guangzhou) and manbf.com (Chongqing, now manbf.net). These sites were chosen for their hosts' geographic diversity and widespread popularity among MSM across China. Both websites are run by community-based, MSM-focused organizations founded in 1998 and have social networking and microblogging features. They also advertise MSM-geared products and events, and provide information and services to promote HIV/STD education and testing. Both organizations are community partners with the local Chinese Centers for Disease Control and have received support from international donors including the Global Fund to Fight AIDS, Tuberculosis and Malaria. Guangzhou tongzhi (GZTZ), also known as the Lingnan Group, has been a partner of the China-Gates Foundation HIV Prevention Cooperation Program since 2008. In 2010, GZTZ.org recorded over 3.5 million unique visitors. An assessment of GZTZ users in 2010 found that among 1100 participants, the average age was 30 years old, 93% lived in Guangdong Province, 77.9% self-identified as gay, and 80% had attained a college degree [47]. Manbf.net is the website for the Chongqing Tongxin Working Group and had over 10,000 registered users in 2011.

Banner ads were posted on the Web portal home pages (see Multimedia Appendices 1 and 2). Participants who clicked on the banner ad were taken to a welcome page describing the survey. Those who chose to continue completed an informed consent screener. Consented participants were asked three eligibility questions: born biologically male, at least 16 years old (age of consent in China), and had ever had anal sex with men.

To complete the survey via GZTZ.org, participants had to register and log in. A list of user names was kept (unaffiliated with survey responses) to deter men from completing the survey more than once. In keeping with GZTZ's past survey research, survey completers were awarded 500 website credits and 50 site "loyalty points" that can be used to "unlock" additional social media features of the GZTZ website. Following the standard practice of the Chongqing Tongxin Working Group at the time, participants recruited through the manbf website were not required to log in and did not receive incentive points.

Frequencies (Table 1) were conducted to describe sociodemographic characteristics, HIV/STD sexual transmission behaviors, and use of computers and mobile phones. To measure interest in using technology as part of their sexual health care, participants were asked to respond "very interested", "somewhat interested", or "not interested" for each of the following six items: receiving HIV testing reminders by SMS/text message, receiving HIV testing reminders by QQ (the most popular Chinese chat and instant messaging service) or email, sexual health counseling by online chat, receiving sexual health education via websites, receiving sexual health education via mobile phone apps, and receiving sexual health education via Internet/mobile phone games. Participants who reported being "very interested" in one or more of these six questions were categorized as having "interest in eHealth", and this variable was used as the main outcome of interest. This decision was made in order to better understand the most likely users for targeting future eHealth interventions.

Univariate and multivariate associations were calculated using logistic regression. A backward elimination process was used to build multivariate models from the variables listed in the tables, with a P value of <.10 required for retention in the final model. We assessed collinearity using the variance inflation factor. All statistical analyses were conducted using SPSS, version 21.

The Institutional Review Boards of the Guangdong Provincial STD Control Center, the London School of Hygiene and Tropical Medicine, and the University of North Carolina at Chapel Hill approved this study. See Multimedia Appendix 3 for the Checklist for Reporting Results of Internet E-Surveys (CHERRIES).

Table 1. Demographic characteristics of a qualitative and Internet sample of MSM in China.

Demographic and behavioral characteristics	Interviews (N=35), n	Internet survey (N=1342[a]), n (%)
Recruitment site [b]		
Guangzhou	28	1204 (89.72)
Hong Kong	7	–
Chongqing	–	138 (10.28)
Residence at time of survey		
Guangzhou City		799 (59.54)
Guangdong Province[c]		331 (24.66)
Chongqing City		124 (9.24)
Other provinces and Hong Kong[d]		66 (4.92)
Country outside of China/Hong Kong		15 (1.12)
Not reported		7 (0.52)
Official registration permit (hukou)		
Guangzhou City		409 (30.48)
Guangdong Province[c]		485 (36.14)
Chongqing City		119 (8.87)
Other provinces and Hong Kong[d]		309 (23.03)
Country outside of China/Hong Kong		13 (0.97)
Not reported		7 (0.52)
Age, mean 30.6, median 29, range 16-64 (n=1327)		
16-20	3	37 (2.79)
21-30	21	733 (55.24)
31-40	9	440 (33.16)
>40	2	117 (8.82)
Marital status (n=1326)		
Married or engaged	5	228 (17.19)
Single	29	1098 (82.81)
Education (n=1331)		
High school or below	7	202 (15.18)
Some college[e]	14	418 (31.40)
College or above	14	711 (53.42)
Living status (n=1326)		
Lives alone	–	466 (35.14)
Lives with others	–	860 (64.86)
Current employment		
Full time	24	1063 (79.21)
Part time/Retired/unemployed	4	151 (11.25)
Student	7	128 (9.54)
Annual income, all sources (USD) (n=1337)		
<6000	–	374 (27.97)
6000-16,000	–	684 (51.16)
>16,000	–	279 (20.87)

Demographic and behavioral characteristics	Interviews (N=35), n	Internet survey (N=1342[a]), n (%)
Sexual orientation (n=1333)		
Gay	28	974 (73.07)
Bisexual	4	352 (26.41)
Heterosexual	–	7 (0.53)
Current stable sex partner (n=1327)		
Male		590 (44.46)
Female	–	70 (5.28)
No stable sex partner	–	33 (2.49)
HIV testing, lifetime (n=1330)		
Never tested	7	525 (39.47)
Ever tested	28	805 (60.53)
HIV status		
Positive		49 (3.70)
Negative/unknown		1293 (96.30)
Any condomless anal sex with men, past 3 months	–	382/988 (38.66)
Ever had sex with women (vaginal or anal)	–	448/1333 (33.61)
Any condomless anal sex with women, past 3 months	–	159/1333 (11.93)

[a] Sample is 1342 unless otherwise noted.

[b] Interview recruitment: on site in person. Survey recruitment: 2 websites hosted in Guangzhou and Chongqing, but open to participants from any location.

[c] Cities outside of Guangzhou but within Guangdong Province.

[d] Cities outside of Chongqing and Guangdong Province.

[e] Includes current students.

Results

Qualitative Sample Characteristics

Overview

Qualitative interviews were conducted with 35 MSM aged 16-48 who were living in Guangzhou and Hong Kong at the time of study recruitment, but whose hometowns represented nine Chinese provinces and Hong Kong. Seven men had never tested for HIV, 7 men had tested for the first time on the day of their interview, and 21 men had tested multiple times. Twenty-eight men self-identified as gay, 4 as bisexual, and 3 chose not to specify. Most participants were unmarried, currently working, and had at least some education beyond high school.

The Internet as a Primary Source of Information About Sexual Health and HIV/STDs

Technology ownership and use was widespread in both the qualitative and Internet survey samples. In the qualitative sample, all men owned a mobile phone, the majority of which were smart phones (27/35). Overall, men had a positive response for using phones and the Internet for sexual health. The most common theme was the level of convenience provided for finding information and getting tested. During the interviews, all men reported using mobile phones or computers to search for general health information, and just over half of the men described examples of finding HIV/STD information online for themselves or for friends through community organizations' websites and other sources:

> *MSM community organizations' websites are very important, they contain comprehensive information…about symptoms, characteristics, the window period etc. When my friend was infected with gonorrhea, I actually went online to look up information for him.* [B04, gay, multi-time tester, age 27]

Notably, the Internet was the primary source of HIV/STD information for many men:

> *R: Did you know much about HIV or syphilis testing before you looked it up online?*
>
> *I: No. I searched all the information online first…lots of the knowledge I've known until now was from the Internet…Like how to protect yourself. What tools that you can use to protect you and your partner while you are having sex. And HIV information, like today, I searched for organizations to have tests.* [A02, gay, multi-time tester, age 24]

In the qualitative interviews, 15 men explicitly remembered using their mobile phones to look up sexual health information via mobile search engines (Baidu, Yahoo, Google, Weibo) and blogs.

The Internet Facilitated HIV/STD Testing

The Internet facilitated HIV/STD testing by providing information about testing and assistance finding testing facilities; creating safe spaces for men to share testing experiences and information about doctors; exposing men to direct HIV/STD test advertising from community organizations, hospitals and health departments; and providing online appointment scheduling and reminders. For many men, learning about HIV/STD testing happened as they were trying to learn about being gay or connecting with the tongzhi (gay men) community. As men learned about and connected with the tongzhi community online, they were also exposed to HIV/STD information, stories, advertisements, and services targeting MSM:

> *On the Internet I learned tongzhi is a population with high risk. I got the [testing center's] address on its website and called to make an appointment…It was a service organization particularly for tongzhi. I am afraid of the way people would look at me if I go to the normal hospital.* [A18, gay, multi-time tester, age 41]

The Internet also provided opportunities to exchange testing experiences and information about doctors. Men used online recommendations, stories, and reviews to make decisions about testing: "Each place offering testing has its own website. I will look at users' comments. If the majority of them agree that a place is good, then that would be my choice" (A17, bisexual, first time tester, age 23).

Men who were connected to the online tongzhi community knew that there were service organizations that would provide free, MSM-tailored testing services and described how they used the Internet as a resource to find these services:

> *If I had a symptom, let's say persistent low fever, then it's possible that I was [HIV] positive, and I'd need to have a test. Instead of going to private doctors where I have to pay, I would probably first go online and look up organizations that provide testing for the gay community. Because they are providing the rapid test for free and have better privacy.* [B07, gay, never tested, age 35]

From a practical standpoint, men used their phones and computers to make testing appointments online, get appointment reminders by text message/SMS, and receive test results by phone call, SMS, or secure website: "Today I used my mobile phone to make my appointment…quite a good experience. Our community organization would inform us when we make an appointment. We would receive text reminders before the appointment telling us the time and duration" (A09, gay, multi-time tester, age 35).

Concerns and Considerations

In spite of widespread use and interest in eHealth for sexual health, men also expressed concerns, most commonly around privacy and confidentiality in the need to protect HIV/STD health information from one's direct social contacts (ie, family, friends, and coworkers) and the general public and government. At the level of privacy from family, friends, and coworkers, the primary concern was that others would directly see HIV/STD information on participants' phones or computers. This concern also applied to the use of HIV/STD-related mobile phone apps: "I would be a little bit worried about [this information] being on my mobile phone. After all, I'm afraid that somebody will see, and worry about leaks of privacy" (A26, bisexual, multi-time tester, age 48) and "I've actually seen some apps [for HIV]…I find them interesting but I will not download them…I sometimes lend my phone to friends for fun. Therefore I would worry they would see those apps" (B01, gay, multi-time tester, age 23).

Participants who already used their computers and phones to search for sexual health information described the protective measures they took including clearing search history, deleting messages, and using phone and Internet passwords. A further recommendation was that future services should use discrete language: "Some people might fear that the messages would be seen by others. That language… could be organized in a way that protects the information, so that other people can't tell what it's about" (A20, bisexual, first time tester, age 21).

Men were also concerned about protecting their HIV/STD health information from the general public and the government. In these conversations, phones could help or hinder privacy:

> *There isn't much to say about the confidentiality of text messages in China. As you know, all your emails and phone calls can be read or heard by others.* [A12, gay, multi-time tester, age 28]
>
> *Actually, it's the security of hospital data worth worrying about. In mainland China we have done a poor job of keeping the CDC and hospital data private. Such institutions call patients' names when it's ones' turn to get tested – it's a big risk for leaking the name of infected patients to the public. Even if the information is not leaked further, it's still embarrassing. So, the mobile phone security does a better job since only the CDC staff can phone you.* [A07, gay, multi-time tester, age 33]

Beyond privacy protection, many men also worried about the trustworthiness of online information: "I believe 60% of the websites, which means I suspect 40%. I think a formal website has higher credibility" (A04, bisexual, multi-time tester, age 21). As this man describes, formal sites were perceived as more trustworthy, including familiar MSM organizations and government-supported organizations: "You can search online at China Red Cross and you can search some very official sites like the Guangdong Provincial Health Department. If it's got 'Province' or 'City' in the title then it's official, it can be trusted" (A02, gay, multi-time tester, age 24).

Internet Survey Sample

For the Internet survey, 3378 participants clicked on the initial link to read about the survey. Among these, 60.49% (2044/3378) consented to participate in the survey. Within these 2044, 6.60% (135/2044) were not eligible due to being born female (5.2%, 7/135), being under the age of 16 (8.1%, 11/135), or reporting never having had sex with a man (86.7%, 117/135). Of the remaining 1909 eligible participants, 1342 (70.30%) finished the survey.

At the time of taking the survey, 59.54% (799/1342) of participants were living in Guangzhou City and 9.24% (124/1342) were living in Chongqing City. Regarding registered residence permit status at the time of the survey: 30.48% (409/1342) were registered in Guangzhou City and 8.87% (119/1342) were registered in Chongqing City (Table 1). For the Internet sample overall, average age was 30.6 years old (range 16-64). Over half of the sample was between 21 and 30 years old. Although the majority of men were single, 64.86% (860/1326) reported living with others (roommates, girlfriend, boyfriend, coworkers, family). Most men had completed an education of college or beyond, and 9.54% (128/1342) were currently full-time students. About three quarters of the sample self-identified as tongzhi (gay). Annual income varied with 27.97% (374/1337) earning less than US $6000 per year. Over one-third of men had never tested for HIV. Among those who had tested and knew their results, 6.1% (49/805) were HIV-positive. A substantial proportion of men reported having condomlesss sex with men (38.7%, 382/988) or women (11.93%, 159/1333) in the past 3 months (Table 1) including 39% (19/49) of men who reported being HIV-positive.

All of our qualitative findings were confirmed in the Internet survey (Table 2): 92.32% (1239/1342) of survey respondents owned a mobile phone or a smartphone and 94.14% (120/1296) owned a computer. The majority (88.05%, 1009/1146) of men connected to the Internet at least once a day and 83.71% (956/1142) used mobile phone apps at least once a day. In the online survey, men reported using computers and mobile phones to search online for general health information (computer: 63.65%, 746/1172; mobile phone: 41.48%, 465/1121) as well as specific information about HIV/STDs (computer: 51.80%, 603/1164; mobile phone: 27.64%, 306/1107).

Most men expressed interest in using eHealth modalities for sexual health (Table 2), including receiving HIV testing reminders by SMS/text message, or QQ chat/email; sexual health counseling by online chat; and sexual education via websites, mobile phone apps, or Internet/mobile phone games. Overall 43.89% (589/1342) of men were very interested in at least one of these eHealth modalities (main outcome), and 29.06% (390/1342) were somewhat or very interested in all six items.

Table 2. Computer and mobile phone usage and acceptability of technology-based sexual eHealth intervention among 1342 MSM in China (N=1342[a]).

Characteristics	n (%)
Technology ownership	
Mobile phone	1239/1342 (92.32)
Smartphone	1146/1225 (93.55)
Computer	1220/1296 (94.14)
Phone service carrier (n=1216)	
CMCC	773 (63.57)
China Unicom	273 (22.45)
China Telecom	149 (12.25)
Other	21 (1.73)
Years using current phone number (n=1226)	
Less than 3 years	514 (41.92)
3 or more years	712 (58.08)
Main study outcome: Very interested in one or more of the following	
Receiving HIV testing reminders by SMS or text message	247/1159 (21.31)
Receiving HIV testing reminders by QQ chat or email	241/1139 (21.16)
Sexual health counseling by online chat	335/1129 (29.67)
Sexual health education via websites	386/1142 (33.80)
Sexual health education via mobile phone apps	306/1126 (27.18)
Sexual health education via Internet/mobile phone games	262/1106 (23.69)
Use mobile phone apps daily	956/1142 (83.71)
Use Internet daily	1009/1146 (88.05)
Past eHealth use, HIV/STD	
Used mobile phone to find HIV/STD testing site	298/1165 (25.58)
Used mobile phone to find other information about HIV/STD	306/1107 (27.64)
Used computer to find HIV/STD testing site	601/1227 (48.98)
Used computer to find other information about HIV/STD	603/1164 (51.80)
Past eHealth use, general	
Used mobile phone for general health information	465/1121 (41.48)
Used computer for general health information	746/1172 (63.65)
Any confidentiality concerns	
Concerns receiving HIV/STD-related SMS or text message	962/1170 (82.22)
Concerns receiving HIV/STD-related QQ chat or email message	946/1163 (81.34)

[a] The total sample size was 1342. Due to programmed skip patterns and participant choice to refrain from answering some questions, denominators in this table can vary from 1342. For example, all participants responded to the question “Do you own a mobile phone?” (1239/1342). The 1239 participants who responded “yes, I own a mobile phone” were then asked, “Is your mobile phone a smartphone?” Among the 1225 participants who chose to answer this question, 93.55% (1146/1225) reported that their mobile phone is a smartphone.

In bivariate analysis (Table 3), being very interested in using computers or mobile phones for sexual health was positively associated with student status, lower annual income, past HIV testing, being self-reported HIV-positive, daily mobile phone app use, daily Internet use, past use of computers or mobile phones to find HIV/STD information or testing sites, past use of computers or mobile phones to find general health information, and confidentiality concerns about using eHealth for sexual health.

Table 3. Correlates of interest in technology for sexual health among 1342 MSM in China.

Characteristics	Very interested in using technology for sexual health (n=589, 43.89%)				
	n/N (%)	OR (95% CI)	*P*	Adjusted OR (95% CI)	*P*
Age					
Over 40	45/117 (38.46)	1			
31-40	186/440 (42.27)	1.17 (0.77-1.78)	.46		
21-30	333/733 (45.43)	1.33 (0.89-1.99)	.16		
16-20	19/37 (51.35)	1.69 (0.80-3.56)	.17		
Education					
College or above	304/711 (42.76)	1			
Some college	183/418 (43.78)	1.04 (0.82-1.33)	.74		
High school or less	97/202 (48.02)	1.24 (0.90-1.69)	.18		
Annual income, RMB					
16000	124/279 (44.44)	1		1	
6000-16,000 RMB	270/684 (39.47)	0.82 (0.62-1.08)	.16	0.76 (0.54-1.05)	.09
6000 RMB	194/374 (51.87)	1.35 (0.99-1.84)	.06	1.01 (0.67-1.52)	.96
Sexual orientation					
Bisexual/heterosexual/other	151/359 (42.06)	1			
Gay	437/974 (44.87)	1.12 (0.88-1.43)	.36		
Living status					
Lives alone	193/466 (41.42)	1			
Lives with others	390/860 (45.35)	1.17 (0.93-1.47)	.17		
Marital status (to women)					
Single/divorced/widowed	489/1098 (44.54)	1			
Married/engaged	93/228 (40.79)	0.86 (0.64-1.15)	.30		
Employment					
Unemployed/part-time/ retired	57/151 (37.75)	1		1	
Full-time employed	451/1063 (42.43)	1.22 (0.86-1.73)	.28	1.20 (0.76-1.90)	.43
Student	81/128 (63.28)	2.84 (1.75-4.63)	<.001[a]	2.27 (1.24-4.16)	.008[a]
Study site					
Guangzhou	522/1204 (43.36)	1			
Chongqing	67/138 (48.55)	1.23 (0.87-1.75)	.25		
Migrant status					
Guangdong or Sichuan	440/994 (44.27)	1			
Other	145/332 (43.67)	0.98 (0.76-1.25)	.85		
Has a current stable sex partner					
No	276/634 (43.53)	1			
Yes	312/702 (44.44)	1.04 (0.84-1.29)	.74		
Unprotected anal intercourse in past 3 months (men/women)					
No	427/971 (43.98)	1			
Yes	134/306 (43.79)	0.99 (0.77-1.29)	.96		
Ever tested for HIV					
No	193/527 (36.62)	1		1	

Characteristics	Very interested in using technology for sexual health (n=589, 43.89%)				
	n/N (%)	OR (95% CI)	*P*	Adjusted OR (95% CI)	*P*
Yes	396/815 (48.59)	1.64 (1.31-2.05)	<.001[a]	1.15 (0.86-1.52)	.35
HIV status					
Negative/unknown	560/1293 (43.31)	1		1	
HIV-positive	29/49 (59.18)	1.90 (1.06-3.39)	.03[b]	1.33 (0.64-2.74)	.44
Use mobile phone apps daily					
No	72/186 (38.71)	1		1	
Yes	452/956 (47.28)	1.42 (1.03-1.96)	.03[b]	1.29 (0.84-1.98)	.25
Use Internet daily					
No	118/333 (35.44)	1		1	
Yes	471/1009 (46.68)	1.60 (1.23-2.06)	<.001[a]	1.08 (0.67-1.74)	.76
Past eHealth use, HIV/STD [c]					
No	174/627 (27.75)	1		1	
Yes	415/715 (58.04)	3.60 (2.86-4.53)	<.001[a]	2.11 (1.53-2.90)	<.001[a]
Past eHealth use, general health					
No	153/561 (27.27)	1		1	
Yes	436/781 (55.83)	3.37 (2.67-4.26)	<.001[a]	1.70 (1.23-2.33)	.001[a]
Any confidentiality concerns					
No	46/302 (15.23)	1		1	
Yes	543/1040 (52.21)	6.10 (4.35-8.55)	<.001[a]	3.70 (2.47-5.54)	<.001[a]

[a] Significant at *P*<.01.

[b] Significant at *P*<.05.

[c] "Past eHealth use for HIV/STI" is a combined variable based on four variables in Table 2 (any past reported use of mobile phone or computer to search for HIV/STI testing site or other information about HIV/STI).

In multivariate analysis (Table 3), four variables remained significantly associated with interest in eHealth: student status (adjusted OR 2.27, 95% CI 1.24-4.16, *P*=.008), past use of eHealth (mobile phone or computer) to search for HIV/STD information or testing site (adjusted OR 2.11, 95% CI, 1.53-2.90, *P*<.001), past use of eHealth to search for general health information (other than HIV/STDs) (adjusted OR 1.70, 95% CI 1.23-2.33, *P*=.001), and having any confidentiality concerns related to eHealth (adjusted OR 3.70, 95% CI 2.47-5.54, *P*<.001).

Discussion

Principal Findings

The purpose of this mixed-methods study among a total of 1377 MSM across multiple sites in China was to conduct formative work to inform the development of future eHealth interventions. We found high ownership, utilization, and interest in computer and mobile phone–based tools to support sexual health care, showing promise for the expansion of eHealth HIV/STD interventions among MSM in China. In this Internet-based sample, MSM who expressed interest in the use of eHealth as part of their sexual health care were more likely to have previously used mobile phones or computers to find HIV/STD information but also had confidentiality concerns about eHealth. Evidence from our qualitative sample elucidated ways men are already using these tools and how they facilitated connections to HIV/STD testing and health services. eHealth tools address barriers to better HIV/STD prevention, linkage to care, and treatment. In particular, men in our sample described how the tools of eHealth addressed barriers to prevention and care activities for them. For example, the Internet provided a way to find testing sites and schedule appointments (removing logistical barriers), identify free services (removing financial barriers), and reduce fear of testing via reading about other men's testing experiences (removing psychosocial barriers).

Our findings speak directly to a recent study among 404 MSM in China that found a lack of HIV testing was associated with not knowing where to get tested, limited HIV knowledge, low perceived HIV risk, concerns about the confidentiality of HIV testing, not being openly gay, not using the Internet, and not having others who had tested for HIV in their social network [48]. Our qualitative interview findings show how Internet use among MSM in China is already addressing each of these barriers and suggest ways for expanding future eHealth interventions. For most men in our qualitative sample, the

Internet was an equally, if not more important source of information about HIV/STDs and testing sites than close friends or health professionals. Men gleaned information from formal sites (health departments, hospitals, community-based organizations) as well as blogs and posts of other men's experiences, reviews, and recommendations. Online advertisements, information, and direct invitations from MSM community-based organizations successfully prompted some MSM to get HIV/STD testing.

Importantly, interactive multimedia has high appeal across gender, culture, and age groups. eHealth modalities are now being used to deliver and address many aspects of HIV prevention including in diverse settings in China such as Web-based HIV/AIDS education and stigma reduction intervention in rural areas [49] and sex education and awareness raising among urban youth [50]. Community-based MSM organizations in China are already trying these technologies. For example, a mobile phone social and sexual networking app called Blued (similar to Jack'd) has been used to promote HIV testing among MSM in Beijing [51]. In addition, the GZTZ organization has created an online program called "Yigaozhi" (Easy to Tell), which helps HIV-positive MSM with status disclosure and HIV test promotion among their sexual partners [52]. The benefits offered by eHealth include broad reach and accessibility to wide audiences; the ability to create interventions that are brief, flexible, and tailored to individual users; and cost savings through scalability, widespread dissemination, and reduced burden compared to face-to-face interventions.

As highlighted in our qualitative interviews, men were aware of the varying quality of information provided online and had developed strategies for coping with this complication including seeking information from sources they perceived to be more "official" such as the Red Cross or provincial-level websites. Credibility is an important consideration in developing eHealth interventions for MSM in China in terms of choosing where and how information will be delivered. There is also room for future research that assesses the accuracy of health information currently being provided by popular MSM websites in China.

Many men accessed the Internet via their mobile phones to look up sexual health information, make appointments, and receive HIV/STD testing information and/or test results. However, overall lower access of the Internet by mobile phone (as compared to computers) in the context of high smartphone ownership indicates an area for development and expansion in optimizing Web and mobile phone apps to make online information more accessible and user-friendly via mobile phones. MSM websites and online services should be visually optimized for mobile phone access to get the broadest use in real-time. Additional tools for future adaptation and development could include websites that include more interactive features, peer and/or health provider facilitated QQ groups and online forums, expansion of reminder and result notification texting/SMS programs, GPS-enabled apps for locating free HIV/STD testing and condoms (eg, NYC Condom, Philly Condom), and mobile phone apps to support HIV care appointments and medication adherence. For example, most men found out about HIV/STD testing online and many men had used online services to schedule appointments. However, there was often a time lag of months or even years between learning about testing and getting tested. This suggests an opportunity for an eHealth-based tool that would facilitate linkage between awareness of testing and actual testing. This kind of tool could include elements such as online personal risk assessment quizzes and automated or personalized follow-up test scheduling prompts and reminders.

To reach the widest population of MSM, eHealth approaches for HIV/STD testing and care should offer a variety of modalities (calls, text/SMS, Web-based resources, online chats/forums, apps, etc) and services. In our Internet survey, MSM preferences for eHealth modalities varied. Sexual health websites, online chat, and mobile phone apps were the most popular. Furthermore, in the qualitative sample, men described a wide variety of ways that they used online HIV/STD information and services. For successful intervention, the mode of eHealth delivery matters. For example, a 2008 randomized controlled trial designed to reduce HIV risk among MSM in Hong Kong failed to achieve significant changes in any of the intervention outcomes [53]. This intervention was primarily delivered to participants by email and followed an information format. Similarly, a multicomponent sexual education Internet-based intervention found that both the email and discussion board features of the intervention were underutilized [50]. In further support of the importance of delivery mode, Zou et al found that MSM in China who were recruited through instant messaging (QQ), online chat, and mobile phone contact (text, SMS) were significantly more likely to actually go and get tested compared to men who were recruited via email [54].

In general, MSM in our study were willing to use their phones to get HIV/STD information and services, but many were concerned about confidentiality and privacy. In the Internet sample, interest in eHealth was strongly associated with confidentiality concerns. This finding may seem counterintuitive; however, it may also signify an unmeasured confounder or construct such as prior negative experiences with breach of confidential health information within the mainstream medical system or general concern about monitoring of the Internet. Further work is needed to better understand the sources of these concerns and how they might be addressed. In the qualitative sample, participants expressed a range of opinions regarding the security of accessing and retaining health information by Internet and on their phones. While some men were concerned about the confidentiality of health information exchanged by phone, others felt that phones were more confidential than paper records or in-person consultations. Those who were using phones and computers for sexual health suggested a range of solutions to protect confidential information including using indirect language, keeping phones locked, and regularly deleting messages and browser history. Overall, men's confidentiality concerns emphasize the importance of sensitivity to this issue as well as building and clearly communicating security features of future eHealth interventions for MSM in China.

Limitations

This study is subject to several limitations. Internet- and venue-based samples are subject to different kinds of recruitment bias. While we cannot quantify the extent of these biases, a strength of our study lies in its use of both recruitment modes and sample findings that mutually reinforced each other, increasing the reliability of both samples. Two additional questions of greatest interest that address the validity of using our findings for intervention recommendations are (1) Does Internet-based recruitment of Chinese MSM capture an appropriate at-risk population for HIV/STD intervention? and (2) How representative is our sample of the broader population of Internet-using MSM in China? In our online sample, men reported high levels of condomless sex with men and women, and low levels of HIV testing. This finding reinforces previous research among MSM in China that found high levels of sexual risk behaviors among Internet-based samples [55-57]. Furthermore, evidence of heightened risk of HIV has been demonstrated among MSM in China who use the Internet. Among a sample of 77 recently diagnosed HIV-positive MSM, 58.4% had found sex partners via the Internet during the year leading up to their infection [58], and among 307 young migrant MSM, men recruited via the Internet had higher prevalence of HIV as compared to those recruited from venues, peer outreach, or network referrals [59].

The health education promotion activities on the two survey host websites may also create a source of bias whereby our recruited participants had higher willingness to utilize eHealth interventions (as compared to the general population of Internet-using MSM). While this is possible, our study found that in spite of exposure to these websites, 38.7% of survey respondents reported condomless sex in the past 3 months and 39.5% had never been tested for HIV. These continued risk behaviors among MSM who are utilizing websites that promote health activities suggest that there is room for intensified or adapted eHealth intervention activities targeting this population. Furthermore, our study does not specifically address MSM who do not have access to (or choose not to use) the Internet and/or mobile smartphones. While likely a growing minority in China's rapidly expanding technology market, these men also need continued, targeted outreach as they are likely to be overrepresented in the lowest socioeconomic strata of MSM in China [5,57].

An additional limitation of our study was the different completion rates between our two recruitment sites. Among screened, eligible participants, 36.5% of Chongqing participants went on to complete the survey versus 78.6% of Guangzhou participants. However, for potential participants who completed the initial screener questions (sex assigned at birth, current self-identified gender, lifetime unprotected anal intercourse, age), we found no significant difference between sites and between those who went on to take the survey versus those who did not. We suspect the difference in the proportion of those who went on to take the survey was most directly related to the lack of incentive provided for the Chongqing participants. One goal in conducting the study was to be least disruptive to the current norms among these well-established online MSM communities. Therefore, we followed each organization's standard practice for remuneration in keeping with their prior research and site management conventions. Given the minor nature of the incentive provided (no money or physical item directly given), we do not feel that the incentive would have caused a meaningful difference in the characteristics of who completed the survey. In addition, we found no association between recruitment site and our outcome variable. We also conducted a sensitivity analysis including recruitment site in the multivariate model and found no significant differences (data not shown). Nevertheless, the varying survey completion rate is an important finding on its own as future Internet-based surveys that attempt to utilize multiple recruitment platforms will face similar questions of following existing remuneration practices versus imposing other standards. It is also promising to note that even a small, non-monetary incentive can serve as sufficient remuneration for eliciting high online survey participation.

Conclusions

To date, meta-analyzed results of interventions among MSM in China have shown significant changes in HIV knowledge and sexual risk behaviors but have not achieved the hoped-for level of change in HIV incidence [60]. The field of HIV prevention is shifting to combination approaches with greater consideration of social and structural risk factors as well as incorporation of biomedical advances (eg, point-of-care rapid HIV, CD4, and viral load testing; ART treatment as prevention, pre-exposure prophylaxis), alongside traditional individual-level education and behavior. Our study has implications for the new "test and treat" models that are being promoted as part of these comprehensive HIV intervention strategies [61]. eHealth interventions can capitalize on the social networking and support available through the Internet and can play a key role along each step of the HIV testing and care cascade from connecting participants to testing centers to helping HIV-infected persons with daily medication adherence [62]. For individual-level interventions, formative work—as we have reported here—and pilot testing among the study population are critical steps for identifying the best modes and messages to improve the chances of intervention efficacy and acceptability. Although Internet and mobile phones are ubiquitous in China, their full potential has not yet been optimally employed to support health interventions for HIV/STD prevention, testing, linkage, and care. eHealth interventions could capitalize on widespread mobile phone ownership and build on preliminary work that establishes the acceptability and current use of these technologies to support MSM health and sexual health in particular.

Acknowledgments

This research was supported by an NIH Career Development Award (1K01TW008200), an NIH institutional training grant (5T32AI007001-35), the UNC Center for AIDS Research (NIAID 5P30AI050410-13), the Partnership for Social Science Research on HIV/AIDS in China (NICHD 1R24 HD056670-04), a junior faculty award from the American Society of Tropical Medicine and Hygiene/Burroughs Welcome Fund, Columbia University College of Physicians and Surgeons Dean’s International Grant, Harvard Medical School, IDSA Medical Scholars Program, and a grant from the Doris Duke Charitable Foundation to UNC-Chapel Hill. We would also like to thank SESH Global (www.seshglobal.org) for technical support.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey banner advertisement ("Participate in this survey to quickly earn 50 loyalty points. NEW!").

[JPG File, 214KB - jmir_v17i4e100_app1.jpg]

Multimedia Appendix 2

Survey banner advertisement ("Survey click here to attend: Understanding MSM's experience and perspectives about testing and obtaining sex products").

[JPG File, 241KB - jmir_v17i4e100_app2.jpg]

Multimedia Appendix 3

Checklist for Reporting Results of Internet E-Surveys (CHERRIES).

[PDF File (Adobe PDF File), 94KB - jmir_v17i4e100_app3.pdf]

References

1. Chen M, Rhodes PH, Hall IH, Kilmarx PH, Branson BM, Valleroy LA, Centers for Disease Control and Prevention (CDC). Prevalence of undiagnosed HIV infection among persons aged ≥13 years--National HIV Surveillance System, United States, 2005-2008. MMWR Morb Mortal Wkly Rep 2012 Jun 15;61 Suppl:57-64 [FREE Full text] [Medline: 22695465]
2. Saxton PJW, Dickson NP, Griffiths R, Hughes AJ, Rowden J. Actual and undiagnosed HIV prevalence in a community sample of men who have sex with men in Auckland, New Zealand. BMC Public Health 2012;12:92 [FREE Full text] [doi: 10.1186/1471-2458-12-92] [Medline: 22296737]
3. Wirtz AL, Jumbe V, Trapence G, Kamba D, Umar E, Ketende S, et al. HIV among men who have sex with men in Malawi: elucidating HIV prevalence and correlates of infection to inform HIV prevention. J Int AIDS Soc 2013;16 Suppl 3:18742 [FREE Full text] [Medline: 24321110]
4. 2012 China AIDS Response Progress Report. 2012. URL: http://www.unaids.org/en/dataanalysis/knowyourresponse/countryprogressreports/2012countries/ce_CN_Narrative_Report%5B1%5D.pdf [accessed 2014-02-17] [WebCite Cache ID 6NSfowGWC]
5. Wu Z, Xu J, Liu E, Mao Y, Xiao Y, Sun X, National MSM Survey Group. HIV and syphilis prevalence among men who have sex with men: a cross-sectional survey of 61 cities in China. Clin Infect Dis 2013 Jul;57(2):298-309 [FREE Full text] [doi: 10.1093/cid/cit210] [Medline: 23580732]
6. Zou H, Hu N, Xin Q, Beck J. HIV testing among men who have sex with men in China: a systematic review and meta-analysis. AIDS Behav 2012 Oct;16(7):1717-1728. [doi: 10.1007/s10461-012-0225-y] [Medline: 22677975]
7. Li X, Lu H, Raymond HF, Sun Y, Jia Y, He X, et al. Untested and undiagnosed: barriers to HIV testing among men who have sex with men, Beijing, China. Sex Transm Infect 2012 Apr;88(3):187-193. [doi: 10.1136/sextrans-2011-050248] [Medline: 22158932]
8. Zhang Y, Lu L, Li HQ, Liu W, Tang ZR, Fang H, et al. Engaging HIV-infected patients in antiretroviral therapy services: CD4 cell count testing after HIV diagnosis from 2005 to 2009 in Yunnan and Guangxi, China. Chin Med J (Engl) 2011 May;124(10):1488-1492. [Medline: 21740803]
9. Lambdin B, Cai T, Udoh I. Identifying bottlenecks: loss to follow-up of MSM from HIV testing to treatment in Wuhan, China. USA: Abstract #LBPE51; 2012 Presented at: XIX International AIDS Conference; July 22-27, 2012; Washington, DC URL: http://pag.aids2012.org/Abstracts.aspx?AID=21557
10. Thompson MA, Aberg JA, Hoy JF, Telenti A, Benson C, Cahn P, et al. Antiretroviral treatment of adult HIV infection: 2012 recommendations of the International Antiviral Society-USA panel. JAMA 2012 Jul 25;308(4):387-402. [doi: 10.1001/jama.2012.7961] [Medline: 22820792]
11. Moore DM, Hogg RS, Yip B, Wood E, Tyndall M, Braitstein P, et al. Discordant immunologic and virologic responses to highly active antiretroviral therapy are associated with increased mortality and poor adherence to therapy. J Acquir Immune Defic Syndr 2005 Nov 1;40(3):288-293. [Medline: 16249702]

12. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
13. Cohen MS, Muessig KE, Smith MK, Powers KA, Kashuba AD. Antiviral agents and HIV prevention: controversies, conflicts, and consensus. AIDS 2012 Aug 24;26(13):1585-1598 [FREE Full text] [doi: 10.1097/QAD.0b013e3283543e83] [Medline: 22507927]
14. Del RJ, Castilla J, Hernando V, Rodríguez C, García S. Combined antiretroviral treatment and heterosexual transmission of HIV-1: cross sectional and prospective cohort study. BMJ 2010;340:c2205 [FREE Full text] [Medline: 20472675]
15. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med 2000 Mar 30;342(13):921-929. [doi: 10.1056/NEJM200003303421303] [Medline: 10738050]
16. Attia S, Egger M, Müller M, Zwahlen M, Low N. Sexual transmission of HIV according to viral load and antiretroviral therapy: systematic review and meta-analysis. AIDS 2009 Jul 17;23(11):1397-1404. [doi: 10.1097/QAD.0b013e32832b7dca] [Medline: 19381076]
17. Lim MSC, Hocking JS, Hellard ME, Aitken CK. SMS STI: a review of the uses of mobile phone text messaging in sexual health. Int J STD AIDS 2008 May;19(5):287-290. [doi: 10.1258/ijsa.2007.007264] [Medline: 18482956]
18. Mukund BKC, Murray PJ. Cell phone short messaging service (SMS) for HIV/AIDS in South Africa: a literature review. Stud Health Technol Inform 2010;160(Pt 1):530-534. [Medline: 20841743]
19. Ybarra ML, Bull SS. Current trends in Internet- and cell phone-based HIV prevention and intervention programs. Curr HIV/AIDS Rep 2007 Dec;4(4):201-207. [Medline: 18366952]
20. Cornelius JB, Cato M, Lawrence JS, Boyer CB, Lightfoot M. Development and pretesting multimedia HIV-prevention text messages for mobile cell phone delivery. J Assoc Nurses AIDS Care 2011;22(5):407-413 [FREE Full text] [doi: 10.1016/j.jana.2010.11.007] [Medline: 21256053]
21. Cornelius JB, Cato MG, Toth JL, Bard PM, Moore MW, White A. Following the trail of an HIV-prevention Web site enhanced for mobile cell phone text messaging delivery. J Assoc Nurses AIDS Care 2012;23(3):255-259 [FREE Full text] [doi: 10.1016/j.jana.2011.03.002] [Medline: 21550826]
22. Lim EJ, Haar J, Morgan J. Can text messaging results reduce time to treatment of Chlamydia trachomatis? Sex Transm Infect 2008 Dec;84(7):563-564. [doi: 10.1136/sti.2008.032441] [Medline: 18723584]
23. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
24. Hardy H, Kumar V, Doros G, Farmer E, Drainoni ML, Rybin D, et al. Randomized controlled trial of a personalized cellular phone reminder system to enhance adherence to antiretroviral therapy. AIDS Patient Care STDS 2011 Mar;25(3):153-161 [FREE Full text] [doi: 10.1089/apc.2010.0006] [Medline: 21323532]
25. Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database Syst Rev 2012;3:CD009756. [doi: 10.1002/14651858.CD009756] [Medline: 22419345]
26. da Costa TM, Barbosa BJP, Gomes e Costa Durval A, Sigulem D, de Fátima MH, Filho AC, et al. Results of a randomized controlled trial to assess the effects of a mobile SMS-based intervention on treatment adherence in HIV/AIDS-infected Brazilian women and impressions and satisfaction with respect to incoming messages. Int J Med Inform 2012 Apr;81(4):257-269 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.10.002] [Medline: 22296762]
27. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de WD, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 2011 Mar 27;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]
28. Dowshen N, Kuhns LM, Johnson A, Holoyda BJ, Garofalo R. Improving adherence to antiretroviral therapy for youth living with HIV/AIDS: a pilot study using personalized, interactive, daily text message reminders. J Med Internet Res 2012;14(2):e51 [FREE Full text] [doi: 10.2196/jmir.2015] [Medline: 22481246]
29. Swendeman D, Rotheram-Borus MJ. Innovation in sexually transmitted disease and HIV prevention: internet and mobile phone delivery vehicles for global diffusion. Curr Opin Psychiatry 2010 Mar;23(2):139-144 [FREE Full text] [doi: 10.1097/YCO.0b013e328336656a] [Medline: 20087189]
30. Juzang I, Fortune T, Black S, Wright E, Bull S. A pilot programme using mobile phones for HIV prevention. J Telemed Telecare 2011;17(3):150-153. [doi: 10.1258/jtt.2010.091107] [Medline: 21270049]
31. Hightow-Weidman LB, Pike E, Fowler B, Matthews DM, Kibe J, McCoy R, et al. HealthMpowerment.org: feasibility and acceptability of delivering an internet intervention to young Black men who have sex with men. AIDS Care 2012;24(7):910-920 [FREE Full text] [doi: 10.1080/09540121.2011.647677] [Medline: 22272759]
32. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. AIDS Patient Care STDS 2013 Apr;27(4):211-222 [FREE Full text] [doi: 10.1089/apc.2012.0404] [Medline: 23565925]

33. Ramanathan N, Swendeman D, Comulada WS, Estrin D, Rotheram-Borus MJ. Identifying preferences for mobile health applications for self-monitoring and self-management: focus group findings from HIV-positive persons and young mothers. Int J Med Inform 2013 Apr;82(4):e38-e46 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.05.009] [Medline: 22704234]
34. 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]
35. Cell phones approach total penetration globally with smartphones moving toward market dominance. 2011. URL: http://www.prnewswire.com/news-releases/cell-phones-approach-total-penetration-globally-with-smartphones-moving-toward-market-dominance-133877063.html [accessed 2014-02-17] [WebCite Cache ID 6NSg7wA67]
36. Li Q, Liu Y, Zhou Z, Li S, Luo F, Li D, et al. Online sex-seeking behaviors among men who have sex with men: implications for investigation and intervention. AIDS Behav 2012 Aug;16(6):1690-1698. [doi: 10.1007/s10461-011-0005-0] [Medline: 21785872]
37. Zhang D, Bi P, Lv F, Zhang J, Hiller JE. Differences between Internet and community samples of MSM: implications for behavioral surveillance among MSM in China. AIDS Care 2008 Oct;20(9):1128-1137. [doi: 10.1080/09540120701842829] [Medline: 18825519]
38. Nehl EJ, He N, Wang X, Lin L, Wong FY, Yu F. Feasibility and willingness of using e-technologies for HIV prevention and research targeting Chinese MSM. AIDS Care 2013;25(7):926-930 [FREE Full text] [doi: 10.1080/09540121.2012.726344] [Medline: 23061806]
39. Tucker JD, Chen XS, Peeling RW. Syphilis and social upheaval in China. N Engl J Med 2010 May 6;362(18):1658-1661. [doi: 10.1056/NEJMp0911149] [Medline: 20445179]
40. Chen ZQ, Zhang GC, Gong XD, Lin C, Gao X, Liang GJ, et al. Syphilis in China: results of a national surveillance programme. Lancet 2007 Jan 13;369(9556):132-138. [doi: 10.1016/S0140-6736(07)60074-9] [Medline: 17223476]
41. Chen XS, Peeling RW, Yin YP, Mabey DC. The epidemic of sexually transmitted infections in China: implications for control and future perspectives. BMC Med 2011;9:111 [FREE Full text] [doi: 10.1186/1741-7015-9-111] [Medline: 21975019]
42. Zhao J, Cai WD, Gan YX, Zhang Y, Yang ZR, Cheng JQ, et al. A comparison of HIV infection and related risk factors between money boys and noncommercial men who have sex with men in Shenzhen, China. Sex Transm Dis 2012 Dec;39(12):942-948. [doi: 10.1097/OLQ.0b013e31826f356f] [Medline: 23191947]
43. Zhong F, Lin P, Xu H, Wang Y, Wang M, He Q, et al. Possible increase in HIV and syphilis prevalence among men who have sex with men in Guangzhou, China: results from a respondent-driven sampling survey. AIDS Behav 2011 Jul;15(5):1058-1066. [doi: 10.1007/s10461-009-9619-x] [Medline: 19826942]
44. Patton MQ. Qualitative Evaluation & Research Methods: Integrating Theory and Practice Fourth Edition. Newbury Park, CA: Sage Publications; 2015.
45. Zhang N, Parish WL, Huang Y, Pan S. Sexual infidelity in China: prevalence and gender-specific correlates. Arch Sex Behav 2012 Aug;41(4):861-873. [doi: 10.1007/s10508-012-9930-x] [Medline: 22544304]
46. Pan S, Parish WL, Huang Y. Clients of female sex workers: a population-based survey of China. J Infect Dis 2011 Dec 1;204 Suppl 5:S1211-S1217 [FREE Full text] [doi: 10.1093/infdis/jir537] [Medline: 22043034]
47. Huang L, Nehl EJ, Lin L, Meng G, Liu Q, Ross MW, et al. Sociodemographic and sexual behavior characteristics of an online MSM sample in Guangdong, China. AIDS Care 2014;26(5):648-652 [FREE Full text] [doi: 10.1080/09540121.2013.844760] [Medline: 24124988]
48. Huang ZJ, He N, Nehl EJ, Zheng T, Smith BD, Zhang J, et al. Social network and other correlates of HIV testing: findings from male sex workers and other MSM in Shanghai, China. AIDS Behav 2012 May;16(4):858-871. [doi: 10.1007/s10461-011-0119-4] [Medline: 22223298]
49. Tian L, Tang S, Cao W, Zhang K, Li V, Detels R. Evaluation of a web-based intervention for improving HIV/AIDS knowledge in rural Yunnan, China. AIDS 2007 Dec;21 Suppl 8:S137-S142. [doi: 10.1097/01.aids.0000304709.02412.3c] [Medline: 18172382]
50. Lou CH, Zhao Q, Gao ES, Shah IH. Can the Internet be used effectively to provide sex education to young people in China? J Adolesc Health 2006 Nov;39(5):720-728. [doi: 10.1016/j.jadohealth.2006.04.003] [Medline: 17046509]
51. Aran I. New gay dating app in China is getting all kinds of action. 2013 Dec 5. Jezebel URL: http://jezebel.com/new-gay-dating-app-in-china-is-getting-all-kinds-of-act-1477168920 [accessed 2014-12-20] [WebCite Cache ID 6Ux3mTtdW]
52. Avery M, Meng G, Mills S. Two internet-based approaches to promoting HIV counselling and testing for MSM in China. Digital Culture & Education 2014;6:216-229 [FREE Full text]
53. Lau JTF, Lau M, Cheung A, Tsui HY. A randomized controlled study to evaluate the efficacy of an Internet-based intervention in reducing HIV risk behaviors among men who have sex with men in Hong Kong. AIDS Care 2008 Aug;20(7):820-828. [doi: 10.1080/09540120701694048] [Medline: 18608057]
54. Zou H, Wu Z, Yu J, Li M, Ablimit M, Li F, et al. Internet-facilitated, voluntary counseling and testing (VCT) clinic-based HIV testing among men who have sex with men in China. PLoS One 2013;8(2):e51919 [FREE Full text] [doi: 10.1371/journal.pone.0051919] [Medline: 23418417]

55. Zhang D, Bi P, Lv F, Tang H, Zhang J, Hiller JE. Internet use and risk behaviours: an online survey of visitors to three gay websites in China. Sex Transm Infect 2007 Dec;83(7):571-576 [FREE Full text] [doi: 10.1136/sti.2007.026138] [Medline: 17971376]
56. Zou H, Wu Z, Yu J, Li M, Ablimit M, Li F, et al. Sexual risk behaviors and HIV infection among men who have sex with men who use the internet in Beijing and Urumqi, China. J Acquir Immune Defic Syndr 2010 Feb;53 Suppl 1:S81-S87 [FREE Full text] [doi: 10.1097/QAI.0b013e3181c7dd2b] [Medline: 20104115]
57. Tsui HY, Lau JTF. Comparison of risk behaviors and socio-cultural profile of men who have sex with men survey respondents recruited via venues and the internet. BMC Public Health 2010;10:232 [FREE Full text] [doi: 10.1186/1471-2458-10-232] [Medline: 20444297]
58. Lee SS, Tam DKP, Mak DWL, Wong KH. Use of the Internet for sex partnership in men who have sex with men before HIV infection. Public Health 2011 Jul;125(7):433-435. [doi: 10.1016/j.puhe.2011.03.010] [Medline: 21733532]
59. Guo Y, Li X, Fang X, Lin X, Song Y, Jiang S, et al. A comparison of four sampling methods among men having sex with men in China: implications for HIV/STD surveillance and prevention. AIDS Care 2011 Nov;23(11):1400-1409 [FREE Full text] [doi: 10.1080/09540121.2011.565029] [Medline: 21711162]
60. Lu H, Liu Y, Dahiya K, Qian HZ, Fan W, Zhang L, et al. Effectiveness of HIV risk reduction interventions among men who have sex with men in China: a systematic review and meta-analysis. PLoS One 2013;8(8):e72747 [FREE Full text] [doi: 10.1371/journal.pone.0072747] [Medline: 24137497]
61. Nachega JB, Uthman OA, del RC, Mugavero MJ, Rees H, Mills EJ. Addressing the Achilles' heel in the HIV care continuum for the success of a test-and-treat strategy to achieve an AIDS-free generation. Clin Infect Dis 2014 Jul;59 Suppl 1:S21-S27. [doi: 10.1093/cid/ciu299] [Medline: 24926028]
62. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A Systematic Review of Recent Smartphone, Internet and Web 2.0 Interventions to Address the HIV Continuum of Care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190. [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]

Abbreviations

AIDS: acquired immune deficiency syndrome
ART: antiretroviral therapy
CD4: cluster of differentiation 4, glycoprotein found on the surface of immune cells
CMCC: China Mobile Communications Corporation
eHealth: electronic Health
GPS: global positioning systems
GZTZ: Guangzhou Tongzhi
HIV: human immunodeficiency virus
MSM: men who have sex with men
QQ: Tencent QQ, a popular instant messaging software service in China
SMS: short message service/text messaging
STD: sexually transmitted disease
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 12.03.14; peer-reviewed by D Zhang, B Sheoran, M Gilbert, Q Li, M Dogan; comments to author 24.07.14; revised version received 19.12.14; accepted 05.02.15; published 21.04.15.

Please cite as:
Muessig KE, Bien CH, Wei C, Lo EJ, Yang M, Tucker JD, Yang L, Meng G, Hightow-Weidman LB
A Mixed-Methods Study on the Acceptability of Using eHealth for HIV Prevention and Sexual Health Care Among Men Who Have Sex With Men in China
J Med Internet Res 2015;17(4):e100
URL: http://www.jmir.org/2015/4/e100/
doi:10.2196/jmir.3370
PMID:25900881

Original Paper

Six-Month Outcomes of a Web-Based Intervention for Users of Amphetamine-Type Stimulants: Randomized Controlled Trial

Robert J Tait[1,2], BSc, PhD; Rebecca McKetin[3], PhD; Frances Kay-Lambkin[4,5], PhD; Bradley Carron-Arthur[2], BPsych (Hons); Anthony Bennett[2], BAppSc; Kylie Bennett[2], BSc, BA (Hons); Helen Christensen[2,6], PhD; Kathleen M Griffiths[2], PhD

[1]National Drug Research Institute, Faculty of Health Sciences, Curtin University, Perth, Australia

[2]National Institute for Mental Health Research, The Australian National University, Canberra, Australia

[3]Centre for Research on Ageing, Health and Wellbeing, The Australian National University, Canberra, Australia

[4]National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia

[5]Centre for Translational Neuroscience and Mental Health, University of Newcastle, Newcastle, Australia

[6]Black Dog Institute, University of New South Wales and Prince of Wales Hospital, Sydney, Australia

Corresponding Author:
Robert J Tait, BSc, PhD
National Drug Research Institute
Faculty of Health Sciences
Curtin University
GPO Box U1987
Perth, 6845
Australia
Phone: 61 8 92661610
Fax: 61 8 92661611
Email: robert.tait@curtin.edu.au

Abstract

Background: The use of amphetamine-type stimulants (ATS) places a large burden on health services.

Objective: The aim was to evaluate the effectiveness of a self-guided Web-based intervention (“breakingtheice”) for ATS users over 6 months via a free-to-access site.

Methods: We conducted a randomized trial comparing a waitlist control with a fully automated intervention containing 3 modules derived from cognitive behavioral therapy and motivation enhancement. The main outcome was self-reported ATS use in the past 3 months assessed at 3- and 6-month follow-ups using the Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST). Secondary outcomes were help-seeking intentions (general help-seeking questionnaire), actual help seeking (actual help-seeking questionnaire), psychological distress (Kessler 10), polydrug use (ASSIST), quality of life (European Health Interview Survey), days out of role, and readiness to change. Follow-up data were evaluated using an intention-to-treat (ITT) analysis with a group by time interaction.

Results: We randomized 160 people (intervention: n=81; control: n=79). At 6 months, 38 of 81 (47%) intervention and 41 of 79 (52%) control participants provided data. ATS scores significantly declined for both groups, but the interaction effect was not significant. There were significant ITT time by group interactions for actual help seeking (rate ratio [RR] 2.16; *d*=0.45) and help-seeking intentions (RR 1.17; *d*=0.32), with help seeking increasing for the intervention group and declining for the control group. There were also significant interactions for days completely (RR 0.50) and partially (RR 0.74) out of role favoring the intervention group. However, 37% (30/81) of the intervention group did not complete even 1 module.

Conclusions: This self-guided Web-based intervention encouraged help seeking associated with ATS use and reduced days out of role, but it did not reduce ATS use. Thus, this program provides a means of engaging with some sections of a difficult-to-reach group to encourage treatment, but a substantial minority remained disengaged.

Trial Registration: Australian and New Zealand Clinical Trials Registry: ACTRN12611000947909; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=343307 (Archived by WebCite at http://www.webcitation.org/6Y0PGGp8q).

(J Med Internet Res 2015;17(4):e105) doi:10.2196/jmir.3778

KEYWORDS

amphetamine-related disorders; Internet; randomized controlled trial; intervention studies; cognitive therapy

Introduction

Methamphetamine and other amphetamine-type stimulants (ATS) increase the concentrations of monoamine neurotransmitters in the synaptic cleft with associated feelings of increased energy, confidence, and euphoria in the user [1]. It is estimated that between 14 and 52 million people used ATS in 2010. This means that the prevalence of ATS use is second only to cannabis of the illicit drugs [2]. ATS can be highly addictive, particularly the more potent formulations (eg, crystalline) and when used via more rapidly absorbed routes (eg, smoking, injecting) [3]. The prevalence of dependence on ATS and the harms resulting from ATS place a considerable burden on health resources, especially in Asia and Oceania, but also in Europe and North America [2,4].

In Australia, it is estimated that 97,000 people are dependent on ATS but few specialist treatment services are available [5]: data suggest that just 16% of nondependent and approximately 30% of dependent methamphetamine users received any treatment for their drug use in the previous year [6,7]. With no pharmacotherapy currently approved for ATS disorders, treatment relies on face-to-face interventions, typically in the form of cognitive behavioral therapy (CBT) or contingency management, which can be extremely resource intensive, preventing their widespread implementation [8,9].

Even where services exist, they are not always accessed by clients. Overall, approximately 24% of those with a substance use disorder used health services for a mental health problem in the previous year, but service use by young people, particularly young males, was much lower (13.2%) [10,11]. A number of potential physical and psychological barriers have been identified that could inhibit utilization of services, including cost, stigma, lack of awareness, and poor access [12,13]. Web-based interventions have the potential to extend the reach of conventional interventions and could overcome many of these impediments [14].

There has been considerable interest in the development and evaluation of Web-based interventions for tobacco use or alcohol consumption, with a recent systematic review summarizing data from studies involving nearly 40,000 smokers [15] and a review of online interventions for alcohol finding 16 studies comprising more than 5600 participants [16]. However, the development of Internet interventions for illicit drug use is at a more formative stage. A meta-analysis of outcomes for computer- and Web-based interventions for cannabis use reported on 10 studies with a total of 4125 participants. Overall the effect size in reducing consumption was small (*g*=0.16) [17]. Limited data are available for Web- and computer-based interventions for users of other illicit drugs, either targeting users of specific drugs (eg, cocaine) [18] or multiple drugs [19-22]. The authors are not aware of any other interventions specifically targeting users of ATS, although we have previously reported outcomes to 3 months [23].

The current study used a randomized design to evaluate the effectiveness of a Web-delivered fully automated intervention for users of ATS against a waitlist comparison group. It was hypothesized that the intervention group would have a greater reduction in their use of ATS 6 months after starting the intervention than the control group. We also examined whether the intervention resulted in improvements on a range of secondary outcomes (detailed subsequently).

Methods

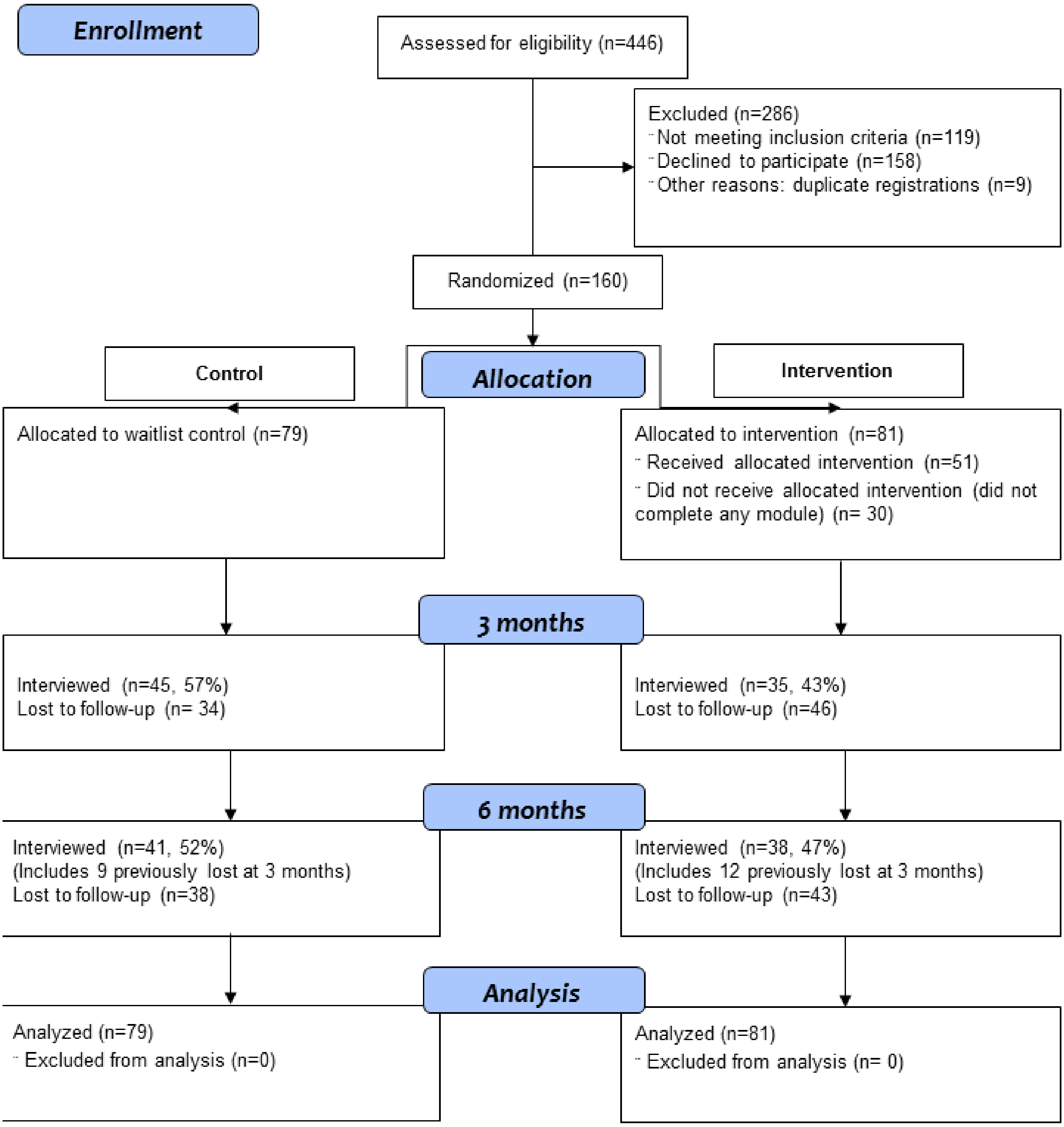
Design

To evaluate the intervention, we used a randomized controlled trial with the intervention group receiving a fully automated, 3-module, Web-delivered intervention. Those in the waitlist group underwent the same assessments as the intervention group, but could not access the intervention for 6 months. Contact details for emergency services were given to participants for crisis support. We have previously described the methodology in detail; the key features are described in the following subsections [24].

Sample

We advertised for participants on social network sites and posters in local clinics; all enrollment processes occurred through the study website. To be eligible, participants had to be resident in Australia, aged 18 years or older, and reported use of ATS (eg, meth/amphetamine, ecstasy, nonmedical use of prescription stimulants) in the previous 3 months. Because of the nature of the intervention, participants were required to have access to the Internet and to provide a valid email address. We excluded potential participants if they reported that they were currently receiving any treatment for stimulant abuse/dependence or methadone, naltrexone, or buprenorphine for a substance use disorder. Those who reported that a doctor had ever diagnosed them as having schizophrenia, schizoaffective, or bipolar disorder were also excluded. Finally, we inspected registration details and 9 cases were excluded as duplicate registrations (eg, identical IP addresses/payment addresses.). Figure 1 provides details of the number recruited and the flow of participants through the study to 6 months. Recruitment ran from January to July 2013.

Figure 1. CONSORT flow diagram for breakingtheice study.



Procedure

All participants were screened and enrolled via the free study website. Those who fulfilled the eligibility criteria were invited to provide consent by "clicking" an onscreen box for each element of the consent form. After consent, the website generated a personalized link that was sent to the participant's email address. This enabled the participants to create their own study username and password. Next, participants were directed to an online baseline survey. Finally, they were randomized using a fully automated system (1-to-1 allocation ratio and with permuted blocks of 4). Participants who were not eligible for the study were provided with information about other potentially relevant websites and community resources.

The intervention group was provided with immediate access to the first module. We recommended that 1 module should be completed each week, but participants were able at advance at their own pace. However, to progress through the program, each page of a module had to be "opened" in sequence to finish that module and progress to the next. Participants could return to any page or module that had already been accessed. Participants received a reminder email 3 days after the expected start date if the first module had not commenced, with a further email sent at day 7 when the next module was due. This pattern of emails was repeated for the other modules. Participants were sent further emails at 3 and 6 months inviting them to complete the follow-up surveys. There was also opportunity at 6 months to provide feedback on the intervention. Participants received AU $20 for each baseline and follow-up assessment, with payment by either posted or online vouchers. The Australian National University Human Research Ethics committee approved the study and it was registered with the Australian and New Zealand Clinical Trials Registry (#12611000947909).

Measures

Outcome data were collected at 3 and 6 months and were all self-reported. The study's primary outcome measure was ATS use assessed with the Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST) [25]. We also collected secondary outcomes on (1) help-seeking intentions (General Help-Seeking Questionnaire) [26], (2) actual help seeking (Actual Help-Seeking Questionnaire) [27,28], (3) psychological distress (Kessler Psychological Distress Scale [K-10]) [29], (4) polydrug use measured by the ASSIST [25], (5) quality of life (European Health Interview Survey Quality of Life scale [EUROHIS]) [30], (6) days partial or wholly out of role [31], and (7) readiness to change (Readiness to Change Questionnaire [RTCQ]) [32]. Demographic details (eg, age, sex), history of drug use (eg, age of first use of ATS), and severity of dependence as measured by the Severity of Dependence Scale (SDS) [33] were collected as part of the baseline survey. The feedback survey included free-text fields plus the 16-item Internet Intervention Adherence Questionnaire [34] and the 16-item Satisfaction with Service measure adapted from the ANU Wellbeing study to reference ATS use rather than depression [35].

Scoring and Coding

The ASSIST appraises lifetime and past 3-month use of 9 classes of drugs (ie, tobacco, alcohol, cannabis, cocaine, ATS, inhalants, sedatives, hallucinogens, opioids, and other). The standard ASSIST scoring algorithm was used to calculate a score for ATS use in the past 3 months, providing scores in the range 0-39 [25] (see Multimedia Appendix 1 for further information). We assessed help-seeking intentions with the question: "How likely is it that you would seek help from each of the following people for any amphetamine or other drug use problems during the next 4 weeks?" For each of 9 potential sources of help, there was a 7-point scale ("extremely unlikely" to "extremely likely") with a score range between 9 and 63. The list did not include breakingtheice. The actual help-seeking measure asked "Which of the following people have you gone to for advice or help in the past 2 weeks for any amphetamine or other drug use problems?" It then listed the same 9 sources of help (range 0-9). We used the total score on the K-10 to evaluate psychological distress. The K-10 uses a 5-point scale ("none of the time" to "all of the time") with a score range between 10 and 50 [29]. To quantify the extent of polydrug use, we summed the ASSIST classes of drugs, excluding ATS use [25]. The total score on the EUROHIS (range 8-40) was used to assess quality of life: higher scores indicate better outcomes [30]. We assessed days completely and partially out of role (both range 0-30) in the previous month using the wording from Kessler's days out of role measure, but referencing "ATS drug use" rather than "depression" [31]. We modified the RTCQ to reference ATS as opposed to alcohol consumption. The measure has 4 items addressing the precontemplation, contemplation, and action stages. Participants were allocated to their highest scoring stage or, in the event of tied scores, to the higher stage [32].

Sample Size

We determined the sample size required to evaluate the primary outcome (ATS score) at a power of 0.8 to detect a medium effect size (eg, d=0.5) [36]. This required 60 people per group, but allowing for 20% attrition, we recruited a total of 80 people per group. The effect size was based on the development study for the ASSIST in Australia [25] and a brief motivational intervention for non-treatment-seeking users of ecstasy [37].

Modules

The content of the modules has been previously described [24]. We based the intervention on principles from motivational interviewing (MI) and CBT and adapted from a face-to-face intervention evaluated in amphetamine users [38]. We also adapted the "decisional balance" approach [39] and asked participants to list both the pros and cons of ATS consumption and the potentially good and bad outcomes anticipated from changing their use of ATS. To help participants change their drug use, we assisted them in the development of skills and strategies to aid in behavioral change (eg, identifying people who could assist them, approaches to help in controlling urges and overcoming cravings, refusal skills, and an action plan to deal with high-risk situations). Example images from the program are available in Multimedia Appendix 2.

Analysis

Initial analyses were conducted in SPSS v21 (IBM Corp, Armonk, NY, USA). Descriptive analyses reported means with standard deviations for continuous measures or percentages for categorical outcomes together with their associated statistics (t test or chi-square test). Effect sizes were calculated as (1) difference in posttest minus pretest means for the 2 conditions divided by their common pretest standard deviation, multiplied by a bias correction factor ($1–3/4[n_{treatment}+n_{control}-2]-1$) [40] and (2) as Cohen's d (posttest intervention mean minus posttest control mean divided by common standard deviation). The characteristics of participants lost to follow-up at 6 months was assessed with logistic regression using baseline predictors of condition, highest education level, age, age of first ATS use, gender, SDS, K-10, ASSIST ATS, polydrug use, RTCQ category, and actual and intended help-seeking scores.

The primary analysis used an intention-to-treat (ITT) approach with the effect of the intervention on each outcome being assessed using a time by group interaction. To analyze the correlated data arising from the repeated measures we used a multilevel mixed-effects regression model with a random intercept term to control for clustering of variance on individuals over repeated measures [41]. This analysis was conducted with Stata SE version 11.2 (StataCorp LP, College Station, TX, USA) using the xtmixed, xtmepoisson, and xtmelogit command suites for linear, Poisson, and logit models, respectively. Measures of days out of role, intended help seeking, and actual help seeking were analyzed using a Poisson distribution. Readiness to change was recoded as a binary variable reflecting action stage versus contemplation or precontemplation stages and analyzed using a logit model. All other outcomes were continuous and analyzed using a linear model. For all measures, we used an unstructured correlation matrix. At baseline, the groups differed significantly on actual help seeking (see Results). To adjust for this difference, baseline actual help seeking was included as a covariate in all models (except for where actual help seeking was the outcome). All models were adjusted for baseline SDS score due to its importance in predicting attrition (see Results).

For the primary outcome (ATS score), we imputed missing data using an iterative Markov chain Monte Carlo (MCMC) method in SPSS to generate 25 sets of data. Maximum and minimum values were logically constrained (eg, to the possible range of scores on the ASSIST), with baseline outcomes and demographic variables used as predictors. The imputed model was conducted in SPSS using the equivalent multilevel mixed-effects linear model to the unimputed model. We also conducted a per-protocol analysis where the “group” variable was replaced with a variable representing exposure to the intervention (“completed any modules,” “completed no modules,” or “control group”).

Results

The majority of participants were male (121/160, 75.6%), the mean age was 22.4 (SD 6.3) years, and 18 of 160 (11.3%) reported using ATS daily or almost daily. In addition, previous treatment for ATS use was reported by 9.4% (15/160) of participants (control: n=7; intervention: n=8) and 23 of 160 (14.4%) reported ever injecting drugs. Table 1 displays the descriptive data at 6 months plus the effect sizes. Baseline characteristics were similar on all measures except for actual help seeking, in which the intervention group had significantly lower levels than the control group (mean 0.3 vs 0.8). (Multimedia Appendix 3 provides mean, SD, and n for each of the outcome variables).

Table 1. Descriptive characteristics by study group at baseline and 6 months plus effect sizes (change from baseline to 6 months on mean scores and between groups at 6 months).

Variable[a]	Baseline		6 months[b]		Effect size, *d*	
	Control n=79	Intervention n=81	Control n=41	Intervention n=38	From 0-6 months	Between groups at 6 months
Sex (male), n (%)	57 (72)[c]	64 (79)				
Age (years), mean (SD)	22.5 (7.1)	22.2 (5.5)				
Age first ATS use, mean (SD)	18.6 (4.2)	17.7 (2.6)				
SDS, mean (SD)	3.8 (3.3)	3.7 (3.5)				
ATS frequency in past 3 months, n (%)						
Never	—	—	8 (20)	5 (13)		
1-2 times	27 (34)	20 (25)	12 (29)	14 (37)		
Monthly	18 (23)	33 (41)	9 (22)	9 (24)		
Weekly	23 (29)	21 (26)	10 (24)	6 (16)		
Daily/almost daily	11 (14)	7 (9)	2 (5)	4 (11)		
ATS score, mean (SD)	16.8 (11.1)	17.0 (10.1)	12.8 (11.1)	13.8 (9.6)	0.07[d]	0.10[d]
Intended help seeking, mean (SD)	20.4 (10.9)	19.7 (11.2)	19.4 (9.2)	22.6 (12.3)	0.32[e]	0.31[e]
Actual help seeking, mean (SD)	0.8 (1.3)[f]	0.3 (0.7)[f]	0.6 (0.9)	0.6 (0.9)	0.45[e]	—
Any actual help seeking (yes), n (%)	34 (43)	20 (25)	15 (37)	16 (42)		
K-10 score, mean (SD)	22.3 (8.3)	22.2 (8.4)	22.0 (8.7)	22.9 (10.0)	0.12[d]	0.10[d]
Polydrug use, mean (SD)	4.6 (1.6)	4.8 (1.8)	4.4 (1.9)	4.5 (2.1)	–0.06[e]	0.05[d]
Quality of life, mean (SD)	28.2 (5.8)	27.2 (6.3)	28.6 (6.8)	27.3 (6.8)	0.05[d]	0.19[d]
Days out of role, mean (SD)	2.9 (5.9)	3.5 (5.6)	2.9 (5.8)	2.8 (6.2)	–0.12[e]	0.02[e]
Days partially out of role, mean (SD)	3.2 (4.8)	3.9 (5.3)	2.8 (4.5)	3.3 (5.7)	–0.04[e]	0.05[d]
RTCQ, n (%)						
Precontemplation	32 (41)	27 (33)	17 (42)	13 (34)		
Contemplation	24 (30)	35 (43)	9 (22)	7 (18)		
Action	23 (29)	19 (24)	15 (37)	18 (47)		

[a] ATS: amphetamine-type stimulants; K-10: Kessler 10; quality of life: EUROHIS score; RTCQ: Readiness to Change Questionnaire; SDS: Severity of dependence.

[b] Missing data at 6 months all outcome variables missing data n=81

[c] One person reported sex as “other”.

[d] Favors control group.

[e] Favors intervention group.

[f] Levene’s correction for inequality of variances (t_{113}=2.83, *P*=.01).

Attrition and Engagement

At 6 months, 41 of 79 (52%) participants from the control and 38 of 81 (47%) from the intervention completed follow-up surveys (Figure 1). Logistic regression showed that retention was not significantly related to group allocation (OR 1.17, 95% CI 0.56-2.47). However, females had higher odds of retention (OR 3.11, 95% CI 1.28-7.55) as did older participants (OR 1.10, 95% CI 1.00-1.20) and those with greater psychological distress (K-10 scores; OR 1.07, 95% CI 1.01-1.14). In addition, higher baseline SDS scores reduced the odds of remaining in the study (OR 0.73, 95% CI 0.59-0.91). In terms of “exposure to the intervention” among the intervention group, 30 of 81 (37%) people did not start or complete the first module, 6 of 81 (7%) completed 1 module only, 6 of 81 (7%) completed 2 modules only, and 39 of 81 (48%) completed all 3 modules. Those who completed any modules (28/51) were not more likely to complete the 6-month follow-up than those who completed zero modules (10/30; χ^2_1=3.5, *P*=.06).

Outcomes

There was a significant main effect on ATS scores, with both groups reducing use by 6 months (b=–2.59, SE 0.98; *P*=.008). However, the interaction term was not significant, showing that the intervention group did not improve more than the control group (Table 2). The ITT analysis was based on those with baseline data plus at least 1 follow-up. There was a significant group by time interaction for actual help seeking (b=0.77, *P*=.02, rate ratio [RR] 2.16, 95% CI 1.14-4.10) and for intended help seeking (RR 1.17, 95% CI 1.05-1.31). There were also significant group by time interactions for number of days out of role (RR 0.50, 95% CI 0.37-0.68) and days partially out of role (RR 0.74, 95% CI 0.56-0.98). In both instances, the intervention group had a greater reduction in days of "impairment" than the control group. Finally, a greater proportion of those receiving breakingtheice transitioned to the action stage than controls (OR 4.13, 95% CI 1.03-16.58). In the analyses involving imputation, the overall group by time interaction for ATS scores was not significant ($F_{1,318}$=0.165, *P*=.69) controlling for baseline SDS and baseline actual help seeking.

Table 2. Statistics for group by time interaction[a] for intention-to-treat (ITT) analyses with unstandardized coefficient (b), standard error (SE), and *P* values.

Variable[b]	Group × 3 and 6 months		Group × 3 months		Group × 6 months	
	b (SE)	*P*	b (SE)	*P*	b (SE)	*P*
ATS score	0.27 (1.41)	.85	0.10 (1.67)	.95	0.87 (1.91)	.65
Intended help seeking	0.16 (0.06)	.005	0.01 (0.07)	.87	0.28 (0.07)	<.001
Actual help seeking	0.77 (0.33)	.02	0.61 (0.39)	.12	0.90 (0.40)	.02
K-10 score	–0.16 (1.20)	.90	–1.79 (1.26)	.16	0.81 (1.71)	.64
Polydrug use	–0.37 (0.30)	.23	–0.64 (0.36)	.08	–0.16 (0.37)	.68
Quality of life	0.52 (0.85)	.55	0.75 (0.95)	.43	0.46 (1.15)	.69
Days out of role	–0.70 (0.16)	<.001	–1.05 (0.24) *p*	<.001	–0.72 (0.18)	<.001
Days partially out of role	–0.30 (0.14)	.04	–0.53 (0.18)	.003	–0.19 (0.17)	.27
RTCQ	1.42 (0.71)	.045	1.04 (0.76)	.17	1.61 (0.93)	.08

[a] Group by time interactions adjusted for actual help seeking at baseline and SDS score at baseline: reference group=control.

[b] ATS: amphetamine-type stimulants; K-10: Kessler 10; quality of life=EUROHIS score; RTCQ: action stage on the Readiness to Change Questionnaire.

Effect of Exposure

The "per-protocol" analyses found significant condition (completed any modules, completed no modules, control group) by time interactions for a number of variables for those who completed any modules compared with controls. For actual help seeking, the rate ratio was 3.13 (95% CI 1.43-6.84) and for intended help seeking it was 1.31 (95% CI 1.16-1.48). Those who completed any modules had significant reductions in both days out of role (RR 0.46, 95% CI 0.33-0.63) and partial days out of role (RR 0.57, 95% CI 0.41-0.77) compared with controls. Those completing any modules were also more likely to transition to the action stage than controls (OR 7.22, 95% CI 1.45-36.01). In one instance, there was a significant effect for those who did not complete any modules compared with controls—those taking no modules had an increase in days partially out of role compared to controls (RR 1.56, 95% CI 1.00-2.42). It should be noted that a per-protocol analysis no longer represents randomized data.

Feedback on the Intervention

Of the 81 people randomized to the intervention, 35 (43%) provided feedback at 6 months. Free-text responses in particular identified the use of fictional case stories as an engaging approach. The main criticisms included the assumption that people wanted to change their behavior and the lack of information on benefits of drug use (eg, use of ATS to control symptoms of attention deficit hyperactivity disorder). The most frequently cited negative reactions to the intervention were concerns about privacy (16/35, 46%) and boredom (7/35, 20%). Most participants (22/35, 63%) reported that using the intervention had reduced their adverse drug effects, 86% (30/35) would recommend the site, 86% (30/35) endorsed Internet delivery, 91% (32/35) rated the site as easy to use, and 91% (32/35) were satisfied with the program.

Discussion

Principal Findings

The results of this study suggest that this fully automated Web-based intervention may be useful both to increase help seeking among people who use ATS and to augment their intention to seek help in the future. There was also evidence for a reduction in the number of days completely and partially out of role. However, the intervention did not reduce ATS use relative to a waitlist control group. Furthermore, relative to the control group, there was no evidence that the intervention reduced the use of other drugs, improved quality of life, or reduced psychological distress.

We believe that breakingtheice is the first Web-based intervention specifically targeting ATS users. However, its lack of impact on ATS use is consistent with 2 previous Internet interventions that included stimulant users. Snow Control targeted cocaine use [18] and found a decline in milligrams of cocaine used per week at 6 months, but no significant effect of the intervention over an attention control group. However, the extremely high rate of attrition at 6 months (94%) makes extrapolation from their findings difficult. A generic online illicit drug (plus alcohol) screening and feedback program developed in Sweden also reported high attrition by 6 months with 69% lost to follow-up. Concordant with the current and Snow Control studies, although drug use declined for both the intervention and controls, there was no significant group by time interaction at 6 months for drug use in the Swedish trial [22].

In contrast, a recent Web-based intervention for cannabis use derived from CBT and MI principles was successful in reducing the frequency of cannabis use, but did not motivate participants to seek additional professional help for their cannabis use, with none of the experimental group seeking additional treatment by 3 months [42]. Although both the Rooke et al [42] and our study involved illicit drug users and adapted treatments from the same paradigms, the profiles of the participants differed with those in the cannabis study being older (mean age 31 years), having higher SDS (mean 14), and younger age of first use (mean 16 years). Further, the reduction in their cannabis use may mean that they did not feel that extra help was required.

The mechanism by which the intervention resulted in fewer days either partially or completely out of role is unclear given that there was not a concomitant decrease in ATS scores or improvements in quality of life or mental well-being. In our sample at baseline, the days out of role equated to 35-42 days per year, which is comparable with international data from high-income countries for drug abuse (mean 37.8 days/year) [43]. At 6 months, there was no change for the controls but the intervention group had fallen to 33.6 days per year, an improvement of more than 8 days per year. Subsequent research is required to identify a plausible mechanism for this change.

Limitations

There are some limitations that should be considered in the interpretation of these results. Firstly, the loss of participants to follow-up threatens the validity and generalizability of conclusions based on these data. However, the rate of attrition (51%) is comparable with the average for fully automated interventions (47%) [44], even though substance users would be regarded typically as a group that is particularly difficult to retain in research projects and treatment. Yet, some study designs have improved follow-up, albeit with a different target group. Interventions that have recruited mother-daughter dyads to prevent cannabis use have achieved at least 90% retention at 12 months [17], but this method requires evaluation in other groups. In addition to loss to follow-up, the fact that a substantial minority in the intervention group did not complete even the first module is of concern, although a low level of engagement has also been identified as a difficulty in face-to-face interventions [45,46]. This low level of engagement may have contributed to the smaller that estimated effect size and the consequent null results. However, the potential for widespread dissemination means that even interventions with small effects can have a public health impact [47].

The representativeness of our sample of ATS users, who were required to have access to the Internet, compared with ATS users in general could be questioned. However, at least in Canada, it appears that users of cocaine or cannabis are as likely to have access to the Internet as current drinkers [48] and the Internet has been shown to be an effective means of reaching hidden populations [49]. Nevertheless, it seems probable that this approach will not reach the most severely disadvantaged ATS users.

Although the feedback on the site was generally positive, we note that these comments only represent a small proportion of the intervention group: we would anticipate that those lost to follow-up would be likely to have more negative opinions. We did not correct for multiple statistical testing, in particular for the secondary outcome measures (eg, all measures other than the ATS score) and, thus, interpret our findings cautiously. Nevertheless, we believe that the inclusion of a range of secondary outcomes is warranted given that this is the first intervention of this type with this population. With the caveat that per-protocol analyses are biased, the per-protocol analyses supports the interpretation that the changes in both help-seeking measures, in days out of role, and transition to the action stage on the RTCQ are not simply type 1 errors due to multiple statistical testing. In each case, it was those who had been exposed to the intervention that showed improvements, whereas those who did not complete at least 1 module showed no improvement compared with those in the control group.

We attempted to prevent duplicate registrations via inspection of IP addresses and payment addresses. This approach does not guarantee that duplicate registrations were eliminated with the potential that dynamic IP addresses or multiple sites could be used together with multiple email or physical addresses.

Implications

The finding that the intervention did not reduce ATS use per se is unsurprising for several reasons. First, most of the intervention focused on enhancing people’s motivation to reduce their ATS use and seek help, with only the later modules focused on strategies for reducing ATS use (module 3). Most of the participants did not complete the last module. Second, the intervention was designed to attract a broad range of ATS users, but a desire to reduce or cease consumption of ATS was not a requirement of the study. With respect to alcohol use, screening and brief interventions in non-treatment-seeking groups have been found to be effective [50]. The majority of participants in this trial were using ATS at low levels (only 11% were using daily or almost daily). Indeed, the feedback on the intervention suggested that most people in the trial were not seeking to reduce their ATS use with less than one-third in the action phase (see Table 1) based on the RTCQ. This may explain why we failed to reduce ATS use. Implementation of the intervention with ATS users who had a greater need or desire for treatment would be required before dismissing its potential to impact on ATS use and related harms. Therefore, it may be necessary to further develop aspects of the module that specifically aim to reduce ATS use.

As noted previously, there was evidence of increased help seeking associated with the intervention, albeit that this was predominantly with informal sources of help. Nevertheless, outside the constraints of a research trial, any engagement with the program could be used as an opportunity to provide information and encouragement to seek further help, particularly for those with low levels of interaction with the online program. There is also the potential to evaluate the intervention as an adjunct to conventional face-to-face treatment. Previous research suggests that compared to a face-to-face CBT intervention alone, online interventions designed to reduce illicit drug use can be effective as an adjunct to weekly individual and group CBT [51]. Integration with face-to-face services could also provide the opportunity to allay privacy concerns expressed by some participants. Furthermore, a Web-based program might allow the extent of face-to-face treatment to be reduced and, hence, lower the burden on service providers and clients inherent in standard treatment.

Conclusions

There is strong evidence for the effectiveness of technological-based programs such as Web- or computer-delivered interventions for problematic use of alcohol or tobacco use [52-55], but their impact on illicit drug use is less certain [14,17]. Nevertheless, this study demonstrates that is possible to engage some ATS users with a Web-based program and retain many of them in a trial to 6 months, but a substantial minority remained disengaged from the process and the effect size across range of measures was small.

Acknowledgments

This study was funded by The Commonwealth of Australia, Department of Health and Ageing. Helen Christensen is funded by NHMRC Fellowship 1056964. Kathleen Griffiths is funded by NHMRC Fellowship 1059620. Frances Kay-Lambkin is funded by NHMRC Fellowship 1008972. Robert Tait is funded by a Curtin University Research Fellowship. None of the funders had any role in study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Scoring the ASSIST.

[PDF File (Adobe PDF File), 28KB - jmir_v17i4e105_app1.pdf]

Multimedia Appendix 2

BTI screenshots.

[PPTX File, 1MB - jmir_v17i4e105_app2.pptx]

Multimedia Appendix 3

CONSORT-EHEALTH checklist V1.6.2 [56].

[PDF File (Adobe PDF File), 80KB - jmir_v17i4e105_app3.pdf]

References

1. Rawson RA, Gonzales R, Ling W. Clinical aspects of methamphetamine. In: Johnson BA, editor. Addiction Medicine: Science and Practice. New York: Springer; 2010.
2. United Nations Office on Drugs and Crime. World Drug Report. Vienna: UNODC; 2012.
3. McKetin R, Kelly E, McLaren J. The relationship between crystalline methamphetamine use and methamphetamine dependence. Drug Alcohol Depend 2006 Dec 1;85(3):198-204. [doi: 10.1016/j.drugalcdep.2006.04.007] [Medline: 16723192]
4. Degenhardt L, Whiteford HA, Ferrari AJ, Baxter AJ, Charlson FJ, Hall WD, et al. Global burden of disease attributable to illicit drug use and dependence: findings from the Global Burden of Disease Study 2010. Lancet 2013 Nov 9;382(9904):1564-1574. [doi: 10.1016/S0140-6736(13)61530-5] [Medline: 23993281]
5. Sara GE, Burgess PM, Harris MG, Malhi GS, Whiteford HA. Stimulant use and stimulant use disorders in Australia: findings from the National Survey of Mental Health and Wellbeing. Med J Aust 2011 Nov 21;195(10):607-609. [Medline: 22107012]
6. McKetin R, Kelly E. Socio-demographic factors associated with methamphetamine treatment contact among dependent methamphetamine users in Sydney, Australia. Drug Alcohol Rev 2007 Mar;26(2):161-168. [doi: 10.1080/09595230601146652] [Medline: 17364851]

7. Wallace C, Galloway T, McKetin R, Kelly E, Leary J. Methamphetamine use, dependence and treatment access in rural and regional North Coast of New South Wales, Australia. Drug Alcohol Rev 2009 Nov;28(6):592-599. [doi: 10.1111/j.1465-3362.2008.00016.x] [Medline: 19930011]
8. Lee NK, Rawson RA. A systematic review of cognitive and behavioural therapies for methamphetamine dependence. Drug Alcohol Rev 2008 May;27(3):309-317. [doi: 10.1080/09595230801919494] [Medline: 18368613]
9. Srisurapanont M, Jarusuraisin N, Kittirattanapaiboon P. Treatment for amphetamine dependence and abuse. Cochrane Database Syst Rev 2001(4):CD003022. [doi: 10.1002/14651858.CD003022] [Medline: 11687171]
10. Reavley NJ, Cvetkovski S, Jorm AF, Lubman DI. Help-seeking for substance use, anxiety and affective disorders among young people: results from the 2007 Australian National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2010 Aug;44(8):729-735. [doi: 10.3109/00048671003705458] [Medline: 20636194]
11. Slade T, Johnston A, Oakley Browne MA, Andrews G, Whiteford H. 2007 National Survey of Mental Health and Wellbeing: methods and key findings. Aust N Z J Psychiatry 2009 Jul;43(7):594-605. [doi: 10.1080/00048670902970882] [Medline: 19530016]
12. Gates P, Copeland J, Swift W, Martin G. Barriers and facilitators to cannabis treatment. Drug Alcohol Rev 2012 May;31(3):311-319. [doi: 10.1111/j.1465-3362.2011.00313.x] [Medline: 21521384]
13. Mojtabai R, Chen L, Kaufmann CN, Crum RM. Comparing barriers to mental health treatment and substance use disorder treatment among individuals with comorbid major depression and substance use disorders. J Subst Abuse Treat 2014 Feb;46(2):268-273 [FREE Full text] [doi: 10.1016/j.jsat.2013.07.012] [Medline: 23992953]
14. Riper H, Tait RJ. Using the Internet for alcohol and drug prevention. In: Miller P, editor. Interventions for Addiction. New York: Elsevier; 2012:855-863.
15. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010(9):CD007078. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
16. Riper H, Blankers M, Hadiwijaya H, Cunningham J, Clarke S, Wiers R, et al. Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: a meta-analysis. PLoS One 2014;9(6):e99912 [FREE Full text] [doi: 10.1371/journal.pone.0099912] [Medline: 24937483]
17. Tait RJ, Spijkerman R, Riper H. Internet and computer based interventions for cannabis use: a meta-analysis. Drug Alcohol Depend 2013 Dec 1;133(2):295-304. [doi: 10.1016/j.drugalcdep.2013.05.012] [Medline: 23747236]
18. Schaub M, Sullivan R, Haug S, Stark L. Web-based cognitive behavioral self-help intervention to reduce cocaine consumption in problematic cocaine users: randomized controlled trial. J Med Internet Res 2012;14(6):e166 [FREE Full text] [doi: 10.2196/jmir.2244] [Medline: 23192752]
19. Ondersma SJ, Svikis DS, Schuster CR. Computer-based brief intervention a randomized trial with postpartum women. Am J Prev Med 2007 Mar;32(3):231-238 [FREE Full text] [doi: 10.1016/j.amepre.2006.11.003] [Medline: 17236741]
20. Schinke SP, Fang L, Cole KC. Computer-delivered, parent-involvement intervention to prevent substance use among adolescent girls. Prev Med 2009 Nov;49(5):429-435 [FREE Full text] [doi: 10.1016/j.ypmed.2009.08.001] [Medline: 19682490]
21. Schwinn TM, Schinke SP, Di NJ. Preventing drug abuse among adolescent girls: outcome data from an internet-based intervention. Prev Sci 2010 Mar;11(1):24-32 [FREE Full text] [doi: 10.1007/s11121-009-0146-9] [Medline: 19728091]
22. Sinadinovic K, Wennberg P, Berman AH. Targeting problematic users of illicit drugs with Internet-based screening and brief intervention: a randomized controlled trial. Drug Alcohol Depend 2012 Nov 1;126(1-2):42-50. [doi: 10.1016/j.drugalcdep.2012.04.016] [Medline: 22613182]
23. Tait RJ, McKetin R, Kay-Lambkin F, Carron-Arthur B, Bennett A, Bennett K. A Web-based intervention for users of amphetamine-type stimulants: 3-month outcomes of a randomized controlled trial. J Med Internet Res Mental Health 2014;1(1):e1-e12.
24. Tait RJ, McKetin R, Kay-Lambkin F, Bennett K, Tam A, Bennett A, et al. Breakingtheice: a protocol for a randomised controlled trial of an internet-based intervention addressing amphetamine-type stimulant use. BMC Psychiatry 2012;12:67 [FREE Full text] [doi: 10.1186/1471-244X-12-67] [Medline: 22731926]
25. Humeniuk R, Dennington V, Ali R, on behalf of the WHO ASSIST Phase III Study Group. The effectiveness of a brief intervention for illicit drugs linked to the alcohol, smoking and substance involvement screening test (ASSIST) in primary health care settings: a technical report of phase III findings of the WHO ASSIST randomized controlled trial. Geneva: World Health Organization; 2008.
26. Wilson C, Deane F, Ciarrochi J, Rickwood D. Measuring help-seeking intentions: Properties of the general help-seeking questionnaire. Can J Couns 2005;39(1):15-28.
27. Rickwood DJ, Braithwaite VA. Social-psychological factors affecting help-seeking for emotional problems. Soc Sci Med 1994 Aug;39(4):563-572. [Medline: 7973856]
28. Rickwood DJ, Deane FP, Wilson CJ, Ciarrochi J. Young people's help-seeking for mental health problems. Australian e-Journal for the Advancement of Mental Health 2005;4(3(Supplement):218-251.
29. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [Medline: 12214795]

30. Schmidt S, Mühlan H, Power M. The EUROHIS-QOL 8-item index: psychometric results of a cross-cultural field study. Eur J Public Health 2006 Aug;16(4):420-428 [FREE Full text] [doi: 10.1093/eurpub/cki155] [Medline: 16141303]
31. Kessler RC, Frank RG. The impact of psychiatric disorders on work loss days. Psychol Med 1997 Jul;27(4):861-873. [Medline: 9234464]
32. Rollnick S, Heather N, Gold R, Hall W. Development of a short 'readiness to change' questionnaire for use in brief, opportunistic interventions among excessive drinkers. Br J Addict 1992 May;87(5):743-754. [Medline: 1591525]
33. Gossop M, Darke S, Griffiths P, Hando J, Powis B, Hall W, et al. The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction 1995 May;90(5):607-614. [Medline: 7795497]
34. Ritterband L. Internet Intervention Adherence Questionnaire. 2011. URL: http://bht.virginia.edu/system/attachments/5/original/Internet%20Intervention%20Adherence%20Questionnaire.pdf?1326484346 [accessed 2012-01-16] [WebCite Cache ID 64j567Qc3]
35. Griffiths KM, Crisp D, Christensen H, Mackinnon AJ, Bennett K. The ANU WellBeing study: a protocol for a quasi-factorial randomised controlled trial of the effectiveness of an Internet support group and an automated Internet intervention for depression. BMC Psychiatry 2010;10:20 [FREE Full text] [doi: 10.1186/1471-244X-10-20] [Medline: 20211025]
36. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
37. Martin G, Copeland J. Brief intervention for regular ecstasy (MDMA) users: Pilot randomized trial of a Check-up model. J Subst Use 2010 Jan;15(2):131-142. [doi: 10.3109/14659890903075074]
38. Baker A, Kay-Lambkin F, Lee NK, Claire M, Jenner L. A Brief Cognitive Behavioural Intervention for Regular Amphetamine Users. Canberra: Australian Government Department of Health and Ageing; 2003.
39. Miller WR, Rollnick S. Motivational Interviewing: Preparing People to Change Addictive Behavior. New York: Guilford Press; 1991.
40. Morris SB. Estimating effect sizes from pretest-posttest-control group designs. Organizational Research Methods 2007 Jul 23;11(2):364-386. [doi: 10.1177/1094428106291059]
41. Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch Gen Psychiatry 2004 Mar;61(3):310-317. [doi: 10.1001/archpsyc.61.3.310] [Medline: 14993119]
42. Rooke S, Copeland J, Norberg M, Hine D, McCambridge J. Effectiveness of a self-guided web-based cannabis treatment program: randomized controlled trial. J Med Internet Res 2013;15(2):e26 [FREE Full text] [doi: 10.2196/jmir.2256] [Medline: 23470329]
43. Alonso J, Petukhova M, Vilagut G, Chatterji S, Heeringa S, Üstün TB, et al. Days out of role due to common physical and mental conditions: results from the WHO World Mental Health surveys. Mol Psychiatry 2011 Dec;16(12):1234-1246 [FREE Full text] [doi: 10.1038/mp.2010.101] [Medline: 20938433]
44. Mathieu E, McGeechan K, Barratt A, Herbert R. Internet-based randomized controlled trials: a systematic review. J Am Med Inform Assoc 2013 May 1;20(3):568-576 [FREE Full text] [doi: 10.1136/amiajnl-2012-001175] [Medline: 23065196]
45. de Weert-van Oene GH, Burger H, Grobbee DE, Schrijvers AJP. Identification of patients at high risk of noncompliance in substance dependence treatment. Eur Addict Res 2007;13(2):74-80. [doi: 10.1159/000097936] [Medline: 17356278]
46. Klein AA, Slaymaker VJ, Dugosh KL, McKay JR. Computerized continuing care support for alcohol and drug dependence: a preliminary analysis of usage and outcomes. J Subst Abuse Treat 2012 Jan;42(1):25-34. [doi: 10.1016/j.jsat.2011.07.002] [Medline: 21862275]
47. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
48. Cunningham JA, Selby PL, Kypri K, Humphreys KN. Access to the Internet among drinkers, smokers and illicit drug users: is it a barrier to the provision of interventions on the World Wide Web? Med Inform Internet Med 2006 Mar;31(1):53-58. [doi: 10.1080/14639230600562816] [Medline: 16754367]
49. Miller PG, Sønderlund AL. Using the internet to research hidden populations of illicit drug users: a review. Addiction 2010 Sep;105(9):1557-1567. [doi: 10.1111/j.1360-0443.2010.02992.x] [Medline: 20626378]
50. Moyer A, Finney JW, Swearingen CE, Vergun P. Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. Addiction 2002 Mar;97(3):279-292. [Medline: 11964101]
51. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Rounsaville BJ. Enduring effects of a computer-assisted training program for cognitive behavioral therapy: a 6-month follow-up of CBT4CBT. Drug Alcohol Depend 2009 Feb 1;100(1-2):178-181 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.09.015] [Medline: 19041197]
52. Myung SK, McDonnell DD, Kazinets G, Seo HG, Moskowitz JM. Effects of Web- and computer-based smoking cessation programs: meta-analysis of randomized controlled trials. Arch Intern Med 2009 May 25;169(10):929-937. [doi: 10.1001/archinternmed.2009.109] [Medline: 19468084]
53. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of E-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011;13(2):e42 [FREE Full text] [doi: 10.2196/jmir.1691] [Medline: 21719411]

54. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]
55. White A, Kavanagh D, Stallman H, Klein B, Kay-Lambkin F, Proudfoot J, et al. Online alcohol interventions: a systematic review. J Med Internet Res 2010;12(5):e62 [FREE Full text] [doi: 10.2196/jmir.1479] [Medline: 21169175]
56. Eysenbach G, Consort- E. 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

ASSIST: Alcohol, Smoking, and Substance Involvement Screening Test
ATS: amphetamine-type stimulants
CBT: cognitive behavioral therapy
ITT: intention-to-treat
MCMC: Markov chain Monte Carlo
MI: motivational interviewing
RTCQ: Readiness to Change Questionnaire
SDS: Severity of Dependence Scale

Edited by G Eysenbach; submitted 12.08.14; peer-reviewed by M Blankers, B Moore; comments to author 14.12.14; revised version received 30.01.15; accepted 19.02.15; published 29.04.15.

Please cite as:
Tait RJ, McKetin R, Kay-Lambkin F, Carron-Arthur B, Bennett A, Bennett K, Christensen H, Griffiths KM
Six-Month Outcomes of a Web-Based Intervention for Users of Amphetamine-Type Stimulants: Randomized Controlled Trial
J Med Internet Res 2015;17(4):e105
URL: http://www.jmir.org/2015/4/e105/
doi:10.2196/jmir.3778
PMID:25925801

Original Paper

Exploring Women's Beliefs and Perceptions About Healthy Eating Blogs: A Qualitative Study

Véronique Bissonnette-Maheux[1], RD, MSc; Veronique Provencher[1], RD, PhD; Annie Lapointe[1], RD, PhD; Marilyn Dugrenier[1]; Audrée-Anne Dumas[1], RD; Pierre Pluye[2], MD, PhD; Sharon Straus[3,4], MD, MSc; Marie-Pierre Gagnon[5,6], PhD; Sophie Desroches[1,5], RD, PhD

[1]Institute of Nutrition and Functional Foods, Laval University, Quebec, QC, Canada
[2]Department of Family Medicine, McGill University, Montreal, QC, Canada
[3]Department of Geriatric Medicine, Toronto University, Toronto, ON, Canada
[4]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada
[5]CHU de Québec Research Center, Population Health and Optimal Health Practices Research Unit, Quebec, QC, Canada
[6]Faculty of Nursing, Laval University, Quebec, QC, Canada

Corresponding Author:
Sophie Desroches, RD, PhD
Institute of Nutrition and Functional Foods
Laval University
2440, Hochelaga boulevard
Pavillon des services, office 2729-P
Quebec, QC, G1V 0A6
Canada
Phone: 1 418 656 2131 ext 5564
Fax: 1 418 656 2131
Email: sophie.desroches@fsaa.ulaval.ca

Abstract

Background: Chronic diseases are the leading cause of death (63%) worldwide. A key behavioral risk factor is unhealthy eating. New strategies must be identified and evaluated to improve dietary habits. Social media, such as blogs, represent a unique opportunity for improving knowledge translation in health care through interactive communication between health consumers and health professionals. Despite the proliferation of food and lifestyle blogs, no research has been devoted to understanding potential blog readers' perceptions of healthy eating blogs written by dietitians.

Objective: To identify women's salient beliefs and perceptions regarding the use of healthy eating blogs written by dietitians promoting the improvement of their dietary habits.

Methods: We conducted a qualitative study with female Internet users living in the Quebec City, QC, area with suboptimal dietary habits. First, the women explored 4 existing healthy eating blogs written in French by qualified dietitians. At a focus group 2-4 weeks later, they were asked to discuss their experience and perceptions. Focus group participants were grouped by age (18-34, 35-54, and 55-75 years) and by their use of social media (users/nonusers). Using a questionnaire based on the Theory of Planned Behavior, participants were asked to identify salient beliefs underlying their attitudes (advantages/disadvantages), subjective norms (what people important to them would think), and perceptions of control (facilitators/barriers) regarding the use of a healthy eating blog written by a dietitian to improve dietary habits. Discussion groups were audiotaped, transcribed verbatim, coded, and a deductive content analysis was performed independently by 2 individuals using the NVivo software (version 10).

Results: All participants (N=33) were Caucasian women aged between 22 to 73 year. Main advantages perceived of using healthy eating blogs written by a dietitian were that they provided useful recipe ideas, improved lifestyle, were a credible source of information, and allowed interaction with a dietitian. Disadvantages included increased time spent on the Internet and guilt if recommendations were not followed. Important people who would approve were family, colleagues, and friends. Important people who could disapprove were family and doctors. Main facilitators were visually attractive blogs, receiving an email notification about new posts, and finding new information on the blog. Main barriers were too much text, advertising on the blog, and lack of time.

Conclusions: The women in this study valued the credibility of healthy eating blogs written by dietitians and the contact with dietitians they provided. Identifying salient beliefs underlying women's perceptions of using such blogs provides an empirically supported basis for the design of knowledge translation interventions to help prevent chronic diseases.

(J Med Internet Res 2015;17(4):e87) doi:10.2196/jmir.3504

KEYWORDS

blog; health behavior; nutrition; qualitative research; social media; knowledge translation

Introduction

Background

Chronic diseases are long-lasting diseases that can be controlled but not cured [1]. They are the leading cause of death (63%) worldwide [2]. In Canada, the proportion of people dying from chronic conditions in 2010 was 89% [2]. The World Health Organization has projected that chronic disease deaths will have increased by 15% globally between 2010 and 2020 [3]. Most countries cannot combat chronic disease with medical care alone: prevention is essential [4]. Healthy eating is recognized as a critical factor in preventing many chronic health conditions, including hypertension, diabetes, cardiovascular diseases, and obesity [5,6]. Food skills, including nutrition knowledge and cooking skills [7], appear to be key to improving eating habits [8] and can help reduce the prevalence of obesity and other chronic diseases in the population [9]. Health communication is an effective strategy for changing norms and beliefs about dietary behaviors, especially by promoting knowledge about appropriate dietary choices [10]. Registered dietitians (RDs) are health care professionals who are trained to transfer this knowledge.

With the advent of Web 2.0, the Web has evolved toward greater simplicity and interactivity. Social media tools are increasingly being used for health communication [11-13], including to deliver health behavior change interventions [14-18], because they reach a large Internet population with diverse sociodemographic characteristics, independently of education, race/ethnicity, or health care access [19]. In the field of physical activity and nutrition, studies have shown promising results [20-26]. A recent literature review published in 2014 that included 22 randomized controlled trials found a significant decrease in dietary fat consumption with the use of social media promoting healthy diet and exercise in the general population [20]. Neuenschwander et al [27] and Bensley et al [28] found that Web-based nutrition education interventions can lead to favorable and effective changes in eating habits when compared with in-person nutrition education interventions. Although these studies show great potential for interventions that use the Internet, none of them included blogs. Blogs are discussion or information sites published by individuals or organizations on the Internet consisting of discrete entries, or posts, displayed in reverse chronological order (the most recent post appears first). Most blogs allow visitors to leave comments and message one another [29].

Studies conducted on health blogs [30-36] and food blogs [37-40] have focused primarily on identifying why bloggers write the blogs and on their information content rather than on blog readers' health behaviors and outcomes. In addition, the majority of these studies refer to food blogs written by lay people and not by RDs. Discerning users have to sift through thousands of blogs to find credible information because noncredible, erroneous, or even harmful health information is widely available and may be increasing eating disorder behaviors [35,36].

To the best of our knowledge, no studies have examined users' perceptions of healthy eating blogs written by RDs. They represent a unique opportunity for improving knowledge translation in nutrition through interactive communication between consumers and dietitians. However, to make use of blog technology to communicate healthy eating information, RDs must better understand how Internet users perceive healthy eating blogs written by qualified nutrition professionals.

Healthy living blogs have been defined as personal webpages devoted to sharing an individual's healthy lifestyle for the purpose of providing an example of "healthy living" to others [35]. There is no standard definition of a healthy eating blog written by an RD; therefore, for the purpose of this study, we defined it as an interactive webpage written by a dietitian promoting the overall improvement of dietary habits (not focused on a specific diet-related disease).

The majority of food bloggers are women [37,39] and their readership is also mostly women who use blogs as an information source about food in general, recipes, or food preparation [41]. Although men also participate in meal preparation, women remain primarily responsible for food purchase and preparation in most households [42,43]. Consequently, health promotion strategies targeting women have the potential to impact not only women blog readers, but also their families.

Purpose

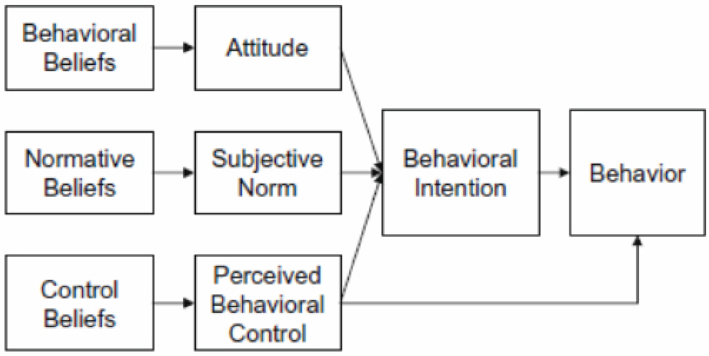
The objective of this qualitative study was to identify the salient beliefs and perceptions of female users/potential users of blogs with suboptimal dietary habits regarding their intention to use a healthy eating blog written by an RD promoting the overall improvement of dietary habits.

Theoretical Framework

A recent systematic review showed that Internet-based interventions developed using a theoretical foundation were more likely to predict behavior than atheoretical interventions [44]. Thus, to maximize the identification of beliefs related to the use of a healthy eating blog written by an RD, we based our focus group moderator's guide and our qualitative analysis on the Theory of Planned Behavior (TPB), which identifies determinants of intention and behavior [45]. We chose this theory because several meta-analytic studies and systematic reviews have shown the strength of its constructs in predicting behavior in a wide variety of contexts [46,47], including healthy eating [48] and, more recently, the use of new Web 2.0 technologies [49-51]. In addition, studies on the adoption of new technologies have suggested a positive correlation between the perception of a blog's usefulness and ease of use and the perceived attitude toward it [52,53].

According to the TPB, intention to use a healthy eating blog written by an RD to improve dietary habits is determined by 3 factors (Figure 1): (1) behavioral beliefs, or underlying attitudes (advantages/disadvantages), toward the use of the blog; (2) normative beliefs, or underlying perceptions of whether important people in one's life would approve or disapprove of the behavior; and (3) control beliefs, or underlying perception of facilitators and barriers to adopting the behavior. Thus, someone will be more inclined to use a healthy eating blog written by an RD to improve his/her dietary habits if he/she perceives more advantages and facilitators than disadvantages and barriers for doing so, and if people who are important to him/her would approve rather than disapprove of the behavior. Developing this knowledge base of beliefs underlying Internet users' attitudes, subjective norms, and perceptions of control regarding healthy eating blogs written by an RD as a means to improve dietary habits will contribute to a better understanding of factors related to the intentions to use and use of these blogs that can contribute to the development of nutritional interventions.

Figure 1. Ajzen's Theory of Planned Behavior [45].



Methods

Participants and Recruitment

We sought female participants with suboptimal dietary habits who were users or potential users of healthy eating blogs. Participants were recruited using the mailing list of the Institute on Nutrition and Functional Foods at Laval University, Quebec, QC, and ads in local newspapers. A total of 57 women responded to our recruitment call, among whom 5 did not meet our inclusion criteria and 19 were no longer interested after they received full information about the nature of their participation. Inclusion criteria were (1) to be a woman living in the Quebec City metropolitan area aged ≥18 years, (2) to have access to the Internet, (3) to use the Internet more than once a week, and (4) to consume 5 or less portions of fruit or vegetables per day (an RD assessed participant's fruit and vegetable consumption over the past 24 hours). Fruit and vegetable consumption is considered a good predictor of overall diet quality [54]. Participants received no honorarium and all gave written informed consent. This project was approved by the Ethics Committee of Laval University (2012-204 A-1/18-03-2013).

Data Collection Procedure

Participants were first interviewed individually and then invited to participate in 1 of 6 focus groups. The focus groups included 4 to 6 women each and were conducted between April and June 2013, 2 to 4 weeks following the individual interview. All focus groups were audiotaped and transcribed verbatim.

During the individual interview, the 33 women were first invited to complete an online sociodemographic questionnaire including 29 questions about their Internet use; the frequency with which they surfed the Internet, read, or commented on blogs; and if they ever looked for nutritional information online. Questionnaires were validated with 8 women to confirm the clarity of the questions prior to this study. Then, semistructured individual interviews were conducted face-to-face with a research coordinator and scheduled to last approximately 1 hour. This face-to-face interview was devoted to the exploration of 4 popular blogs written in French by RDs that promoted healthy eating. The 4 blogs were the same for all the participants and were chosen before the interview. To identify the 4 blogs, we used the Google search engine and typed the following French keywords: blog, *blogue* and *nutritionniste* or *diététiste* (dietitian). Blogs were included if they met the following criteria: (1) written in French by a Canadian RD, (2) targeted human nutrition as the unique topic, (3) had as its main objective the overall promotion of a healthy diet (ie, did not focus on a specific diet-related disease), and (4) was proactive (ie, had published a new post at least once a month since the creation of the blog and had published a minimum of 12 posts). During the interview, participants were asked their perceptions about design features and nutritional content of those 4 blogs to explore different facets characterizing existent healthy eating blogs written by dietitians. Because some women were not frequent users of blogs or other social media, the primary goal of the interview was to prepare the women to discuss their perceptions of healthy eating blogs written by an RD at the subsequent focus group.

Between 2 and 4 weeks after the individual interviews, the 33 women were invited to participate in a focus group of 90 minutes moderated by a trained research coordinator and an assistant

(VBM). The moderator used a semistructured interview to ask participants questions about their perceptions of consulting healthy eating blogs written by RDs to promote improvement of dietary habits, and the assistant took notes during the discussion. The procedure and interview guide were based on Patton’s recommendations [55]. The standardized open-ended interview questionnaire was developed according to the 3 constructs of the TPB. Questions aimed to identify the salient beliefs underlying their attitudes (advantages/disadvantages), subjective norms (approval or disapproval of important others), and perception of control (facilitators/barriers) with regard to the use of blogs written by RDs promoting improvement of dietary habits, such as those they had encountered during the individual interviews. All women in each focus group were asked the same questions in the same order to increase comparability of responses. The interview guide was validated with a focus group of 6 women before conducting the experimental focus groups [55].

Six focus groups (4 to 6 women in each focus group) were needed to achieve theoretical saturation [56]. To determine saturation, we calculated the extent to which different focus groups mentioned the same themes. By the end of the fifth focus group, 98.6% of themes had been mentioned at least once; the remaining 1.4% of themes were only mentioned in the sixth focus group. Participants were grouped by age and by their use of social media (Facebook, Twitter, or blogs) to increase homogeneity and better describe population subgroups as per recommendations by Patton and collaborators [55]. Two focus groups included women aged between 18 and 34 years who used social media, 2 focus groups included women between 35 and 54 years who used social media, and 2 focus groups included women older than 55 years who did not use social media on a regular basis.

Data Analysis

Descriptive statistics and mean ± SD were used to analyze all quantitative data from the online sociodemographic questionnaire using the SAS version 9.3 (SAS Institute, Inc, Cary, NC, USA). The steps of the deductive content analysis described by Elo et al [57] inspired the content analysis of the focus groups, which were transcribed verbatim. Quotes were coded line by line to bring out the main salient beliefs according to the construct of the TPB: behavioral beliefs, normative beliefs, and control beliefs (Figure 1). Two coders (VBM and MD) performed the coding with NVivo version 10 (QSR International, Cambridge, MA, USA) independently and then compared themes to reach consensus on the terminology to be used for each. A third person (AAD) validated the resulting categories and was available to resolve any discrepancies. For the purpose of our study, all beliefs that emerged in at least 2 groups of 6 (33%) were considered salient beliefs [58]. If the same belief was named by several participants in the same group, it was considered a single belief. Finally, categories obtained for groups according to participants’ age were compared to see if any age-related differences in salient beliefs emerged.

Results

Of the 33 women who participated in the individual interviews, 29 also participated in 1 of the focus groups. The 4 women who did not complete the project mentioned lack of time or unforeseen personal circumstances on the date scheduled for the focus group. All participants were Caucasian females between 22 and 73 years of age, most were fairly highly educated and of above-average income (Table 1). Of the 33 women, 25 (76%) had consulted a blog before, but only 3 spent more than 10 hours a week on the Internet (Table 2). Yet the majority identified the Internet as their principal source of information about health, nutrition, and recipes (Figure 2).

Results of the analysis of the 6 focus groups are presented in Table 3, showing the women’s salient beliefs reported in each category and the frequency of mentions according to age group. Select quotes presented subsequently illustrate our key findings. Quotes were originally in French and have been translated by a professional translator.

Table 1. Sociodemographic characteristics of the 33 participants.

Characteristics	Participants
Age (years), mean (SD)	44 (17)
Ethnicity, n (%)	
Caucasian	33 (100)
Highest level of education completed, n (%)	
Primary	1 (3)
High school	3 (9)
College	13 (39)
University	16 (48)
Family income (Can $), n (%)	
0-19,999	4 (12)
20,000-49,999	10 (30)
50,000-99,999	8 (24)
100,000-149,999	4 (12)
150,000-199,999	5 (15)
≥200,000	0 (0)
Unknown	2 (6)

Table 2. Internet use characteristics of the 33 participants.

Internet characteristics	n (%)
Time spent on Internet for leisure	
1-3 hours/week	13 (39)
3-5 hours/week	6 (18)
5-10 hours/week	11 (33)
≥10 hours/week	3 (9)
Read a blog before	25 (76)
Read a blog on nutrition before[a]	16 (64)
Read comments on a blog before[a]	18 (72)
Commented on a blog before[a]	2 (8)

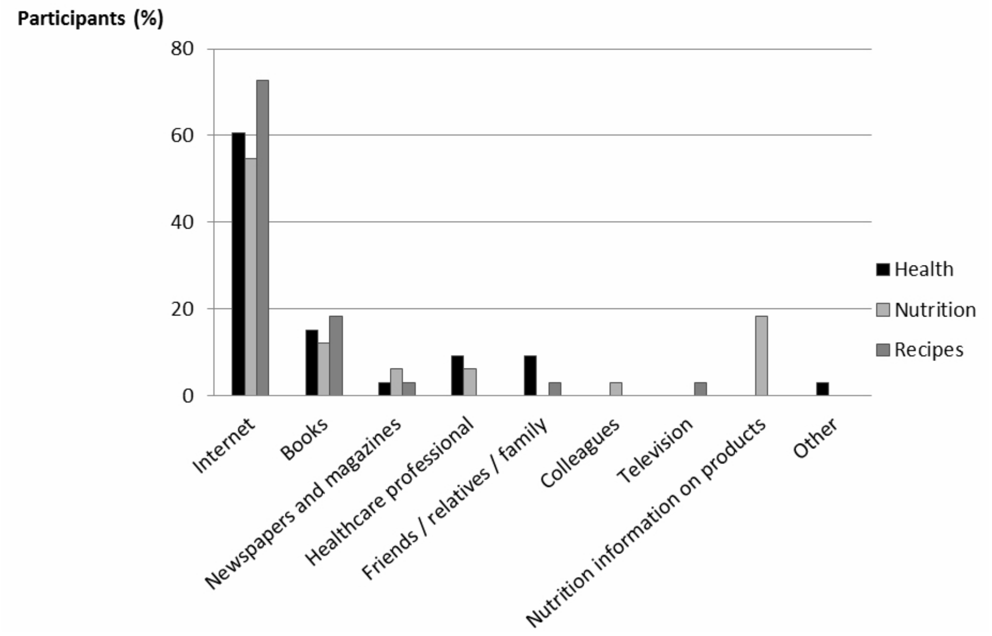
[a] Only among participants who had already read a blog before (n=25).

Table 3. Salient beliefs associated with the use of a healthy eating blog.

Salient beliefs	Frequency of groups, n (n=6)	Frequency of mentions by age (years), n		
		18-34 (n=2)	35-54 (n=2)	55-75 (n=2)
Behavioral beliefs-perceived advantage				
Gives recipe ideas	5	2	2	1
Helps improve diet-related lifestyle	4	1	1	2
Promotes interaction with the dietitian	4	2	0	2
Interesting to have the opinion of fellow blog readers	3	1	2	0
Introduces to new foods	3	1	0	2
Introduces to new trends in nutrition	3	1	1	1
Allows to learn more about nutrition in general	3	0	2	1
Credible information	3	1	0	2
Allows to have the opinion of a dietitian	2	0	1	1
Behavioral beliefs-perceived disadvantages				
Opinion of only 1 blogger (if only 1 blogger is followed)	2	2	0	0
Increases time spent on the Internet	2	1	1	0
Nonrelevant comments from blog readers	2	1	1	0
Feeling of guilt if one does not follow dietary advice given on the blog	2	1	1	0
Normative beliefs-approval				
Family	5	2	1	2
Colleagues	5	2	2	1
Friends	4	2	1	1
Physicians	2	0	0	2
Neighbors	2	0	1	1
Normative beliefs-disapproval				
Family	4	2	2	0
Physicians	3	0	1	2
Friends	2	1	0	1
Colleagues	2	1	1	0
Control beliefs-facilitators				
Visually attractive blog	6	2	2	2
Email to inform of a newly published posts	5	1	2	2
New information weekly (updated weekly)	5	1	2	2
Well-designed for quick access to desired information	4	0	2	2
Interesting and entertaining posts	4	2	2	0
Quick access to the blog	3	2	0	1
Well written, concise info	3	1	1	1
Written by a personality	3	1	1	1
Notification through Facebook to inform of a newly published posts	3	2	1	0
Presence of recipes	3	1	2	0
Contains complete information	3	1	2	0
Popular blog, referred by someone	3	1	1	1

Salient beliefs	Frequency of groups, n (n=6)	Frequency of mentions by age (years), n		
		18-34 (n=2)	35-54 (n=2)	55-75 (n=2)
Text well-structured in paragraphs with subtitles	3	0	2	1
Availability of nutritional value of the recipes	2	1	1	0
Opportunity to read other readers' comments and to link with them	2	1	1	0
Possibility to ask questions, interactivity	2	0	1	1
Easy to understand, not too scientific	2	0	1	1
Narrative approach	2	0	2	0
Posts about novel foods, trends	2	1	1	0
Control beliefs-barriers				
Posts too long to read	5	1	2	2
Presence of advertising	5	1	2	2
Smaller font size	4	1	2	1
Conflict of interests, promotion of commercial products by blogger	4	1	2	1
Lack of new information, too repetitive	4	1	1	2
Lack of time	4	1	1	2
Poorly structured, complicated site	4	1	2	1
Notifications too frequent	4	2	1	1
More negative than positive posts	3	1	1	1
Visually unattractive	3	0	2	1
Computer issues (eg, slow)	2	1	0	1
Information not available, not accessible	2	0	1	1
Credibility: can be written by anyone, not always true	2	1	1	0
If it is not professional, poor writing	2	0	2	0

Figure 2. Information sources for information on health, nutrition, and recipes.



Behavioral Beliefs: Perceived Advantages and Disadvantages

The most frequently cited advantages women perceived to consulting a healthy eating blog by an RD were that it gives recipe ideas (5/6 groups, 83%), helps improve diet-related lifestyle (4/6 groups, 67%), and promotes interaction with a dietitian (4/6 groups, 67%):

> *What's useful with following a blog is if you can comment, you create a link with the nutritionist and you can follow her, and so there's a connection that builds up.* [group #3, participant 21]

Numerous participants aged 35 years and older considered healthy eating blogs as a tool for learning more about nutrition in general (3/4 groups, 75%):

I mean food is such a huge topic, it's impossible to know everything about it, so you can go and look at what you like [on the blog] *and that leads you on to new ideas for trying new things out*. (group #1, participant 3)

The 4 most cited disadvantages were having the opinion of only 1 dietitian-blogger (if only 1 blog was followed) (2/6 groups, 33%), increased time spent on the Internet (2/6 groups, 33%), irrelevant comments from other readers on the blog (2/6 groups, 33%), and guilt arising from not complying with the dietary recommendations on the blog (2/6 groups, 33%):

> *If you're working...sometimes, I don't even have time to stop and eat at work, or whatever. So then, I go look at it [the blog], and there are some great recipes. I don't have time to go grocery shopping, I don't have time to make them. And then, ah, I'm so discouraged* [group #3, participant 20]

Although groups comprised of women ≥55 years mentioned few disadvantages (n=2) of consulting a healthy eating blog by an RD, each was only mentioned in 1 group.

Normative Beliefs: Approval and Disapproval

Participants identified many individuals they believed would approve of them reading a blog to improve their dietary habits. Among them, family (5/6 groups, 83%), colleagues (5/6 groups, 83%), and friends (4/6 groups, 67%) were most frequently mentioned as normative referents. Interestingly, most participants (4/6 groups, 67%) considered reading and interacting with a blog as a personal or a private action, so it was unclear to them (and perhaps irrelevant) whether people would approve or disapprove:

> *The things I go and look at on the Internet, they're also personal. I'm not the type to go talking about everything I do. So it's not likely that I'll know if people around me are against it—I keep it to myself. If I'm interested in nutrition, in going on blogs, that's my business, it's not anyone else's.* [group #5, participant 33]

Almost all groups (5/6 groups, 83%) reported that it was difficult to be against healthy eating considering that it is an inoffensive subject. However, some participants identified their family (4/6 groups, 67%) and physicians (3/6 groups, 50%) as the persons most likely to disapprove of their use of blogs to improve dietary behaviors, perhaps because of erroneous or harmful information that can be found on health blogs:

> *My doctor already warned me, because I've told her twice that I go on it. She said for sure there are some good things, but she said to be careful. She meant that not everything out there is good.* [group #2, participant 2]

Control Beliefs: Facilitators and Barriers

All groups mentioned that visual characteristics of the blog were important facilitators:

> *An attractive site, that helps too. When it's just text with no pictures, it's heavy. I'm very...I like it when it's simple, very visual, that makes me want to go look at it.* [group #3, participant 29]

Most women said that facilitators were receiving an email notifying them of a new post published on the blog (5/6 groups, 83%) and having new content or new information each week (5/6 groups, 83%):

> *It shouldn't always repeat the same thing. You can get tired of it, you know. It has to be a little different ... for example, every week something new that will make us want to go check it out* [group #2, participant 2]

Interesting and entertaining posts as well as well-designed blogs allowing quick access to the desired information were also identified frequently as facilitators (4/6 groups, 67%).

In contrast, the most often reported barriers to consulting a healthy eating blog by an RD were posts being too long (5/6 groups, 83%), presence of advertising (5/6 groups, 83%), bloggers being in conflict of interest by promoting products on their blog (4/6 groups, 67%), and lack of time (4/6 groups, 67%).

Lack of new information, repetitive posts, small font sizes, or badly structured sites were also considered barriers to using a blog by most women (4/6 groups, 67%). Finally, receiving notifications of new posts too often was also identified as a barrier (4/6 groups, 67%):

> *If it's irritating, if someone or something doesn't stop bugging me, I'm gonna go, like, ok, this is junk mail.* [group #3, participant 8]

Discussion

Principal Findings

Our objective in conducting this qualitative study was to better understand women's perceptions toward healthy eating blogs written by RDs as a tool to improve their dietary habits. Identifying the salient beliefs underlying women's attitudes, subjective norms, and perceptions of control have important implications for developing interventions to prevent diet-related chronic diseases, including helping dietitians to design user-friendly and relevant healthy eating blogs to improve dietary habits and to assess their efficacy. Many of our results can be read as clear design guidelines regarding details such as blog length, website structure, email notifications, and visual attractiveness. In addition, we believe the following 5 points are worthy of further discussion.

First, results of our study show that healthy eating blogs written by RDs may be an important nutritional knowledge translation tool for preventing chronic disease. The women perceived numerous advantages, including increased knowledge about new foods, nutrition trends, healthy recipes, and knowledge about nutrition in general, and they identified few disadvantages. This is coherent with findings by Edwards et al [59] who reported that among chronic disease patients, the most important constructs related to interest in use of technology to provide health care remotely were (1) confidence in using it and (2) perceiving greater advantages and fewer disadvantages.

Second, the credibility of the blogger and the information found on blogs were not barriers mentioned frequently by participants in our study. Given that many studies have found that credibility is a concern frequently mentioned by users searching for health information on social media [60,61], this suggests that the fact that the blogs in our study were written by RDs made a difference to the participants’ confidence in them. This suggestion is upheld by Greenberg et al [57] who reported that blogs written by experts are perceived as being more credible than personal blogs. However, the extent to which food blogs not written by accredited health care professionals such as dietitians contain erroneous information remains to be studied. In a recent survey of 679 food bloggers, 87% said they were writing a blog primarily because of their passion for food and only 22% had a work history relating to food (not necessarily dietitians) [62]. The credibility of information found on blogs is thus an important concern not only for health care providers, but also for consumers of healthy eating blogs and suggests that there might be an important role for dietitians in designing interventions that will be of interest to consumers.

Third, an interesting feature of blogs as potential tools for promoting better health is their interactive nature [63]. Many women in our study perceived that interacting with others, including the other blog users but especially with an expert dietitian-blogger, was one of the chief advantages of consulting healthy eating blogs written by RDs. Recent studies have shown that increased interaction, shared information, and peer support are key benefits to using social media for health communication [52,53]. In contexts where the health care system provides limited access to live health care professionals, this feature may be especially attractive to potential users.

Fourth, our results show that healthy eating blogs written by a dietitian may not yet be the media of choice for translating nutritional knowledge to older women. First, although they cited few disadvantages (determinant of attitude), they mentioned several barriers (determinant of perceived control), suggesting that their issues with blogs were more in the domain of perceived capacity. Also, the features they appreciated, such as email notifications about new posts rather than a notification through Facebook for instance, suggest that these women may be more comfortable with email than with more recent interactive communication technologies (eg, Facebook). In our study, the nutrition-related content of the post (eg, recipes, quality of information) seemed more important for young women, whereas the site design (structure of the blog and the posts) seemed more important for older women. According to Chou and collaborators [19], age is the greatest predictor of the use of social media, with a lower penetration in the older population aged 55 years and older. However, studies also suggest that this phenomenon is changing. According to a recent Pew Internet survey, 50% of respondents aged 50-64 years reported using social media in 2012 compared to approximately 15% in 2008 [64]. Designers should include users of all ages in the process of developing different knowledge translation strategies.

Finally, our results raise the theoretical question of the use of the unmodified TPB for examining predictors of intention for using social media, particularly the relevance of the subjective norm construct. Women in our study considered that consulting a healthy eating blog by an RD to improve their dietary habits was a personal and individual action that did not necessarily require approval from others. Several cognitive behavior models that account for technology acceptance have been explored in the quest for a “unified theory” that may provide fruitful avenues for modifying or adding to the TPB in future studies of the experiences of potential users of healthy eating blogs written by RDs with a view to developing effective interventions for chronic disease prevention [51,52].

Limitations

Godin et al [58] suggest that a sample of 25-30 participants is appropriate to highlight salient beliefs of a population, the participants in our study were all Caucasians, mostly educated above high school level (college and university), and had a relatively high family salary income; therefore, they are not fully representative of the general population. Second, this was an experimental and not a natural setting. However, as we wished to reach potential users as well as users, we were obliged to introduce potential users to blogs. In addition, we exercised control over the setting because we were exploring blogs specifically written by dietitians.

Conclusion and Implications for Health Care

As the interest for the use of social media, such as blogs, in the population is growing, there is an urgent need to assess their impact on health outcomes. Research in blogging as a potential tool for knowledge translation in health is in its infancy and much work remains to be done before we can determine its effectiveness for improving healthy eating. Our study identified the perceptions of female Internet users with regard to their use of healthy eating blogs written by RDs. Our results can help design interventions that address attitudes, facilitators, and barriers in developing theory-based dietary behavior change interventions with a view to preventing diet-related chronic diseases.

Acknowledgments

We would like to thank the participants of this study for their time, Mylène Turcotte for moderating the focus groups and for her guidance during the qualitative analysis of the results, and Louisa Blair for editing this paper. We also want to express our gratitude to Knowledge Translation Canada for their financial support.

Authors' Contributions

VBM wrote a first draft of this paper. SD, PP, VP, MPG, and SS designed the study. AL, MD, and AAD contributed to the transcriptions and content analysis of the focus groups with VBM. All authors have reviewed and accepted the manuscript. VBM and AL performed the individual interviews and took field notes during the focus groups.

Conflicts of Interest

None declared.

References

1. The Center for Managing Chronic Disease. What is chronic disease? URL: http://cmcd.sph.umich.edu/what-is-chronic-disease.html [accessed 2015-03-23] [WebCite Cache ID 6Uh5e6dJn]
2. World Health Organization. Noncommunicable diseases country profiles 2011. Geneva: World Health Organization; 2011. URL: http://whqlibdoc.who.int/publications/2011/9789241502283_eng.pdf [accessed 2014-04-22] [WebCite Cache ID 6P1iBRyH6]
3. World Health Organization. Global status report on noncommunicable diseases 2010. Geneva: World Health Organization; 2011. URL: http://www.who.int/nmh/publications/ncd_report_full_en.pdf [accessed 2014-04-22] [WebCite Cache ID 6P1k4XMN0]
4. Elmslie K. Against the Growing Burden of Disease. Ottawa, ON: Public Health Agency of Canada; 2012. URL: http://www.ccgh-csih.ca/assets/Elmslie.pdf [WebCite Cache ID 6Uh4LtjE5]
5. Bantle JP, Wylie-Rosett J, Albright AL, Apovian CM, Clark NG, Franz MJ, et al. Nutrition recommendations and interventions for diabetes: a position statement of the American Diabetes Association. Diabetes Care 2008 Jan;31 Suppl 1:S61-S78. [doi: 10.2337/dc08-S061] [Medline: 18165339]
6. Joint WHO/FAO Expert Consultation on Diet, Nutrition and the Prevention of Chronic Diseases. Diet, nutrition and the prevention of chronic diseases. World Health Organ Tech Rep Ser 2003;916:i-viii, 1. [Medline: 12768890]
7. Region of Waterloo Public Health. Food Skills of Waterloo Region Adults. 2010 Jan. URL: http://chd.region.waterloo.on.ca/en/researchresourcespublications/resources/foodskills.Pdf [accessed 2014-11-12] [WebCite Cache ID 6U2E6SoZE]
8. Shaikh AR, Yaroch AL, Nebeling L, Yeh MC, Resnicow K. Psychosocial predictors of fruit and vegetable consumption in adults a review of the literature. Am J Prev Med 2008 Jun;34(6):535-543. [doi: 10.1016/j.amepre.2007.12.028] [Medline: 18471592]
9. Public Health Agency of Canada. 2012. Curbing childhood obesity: A federal, provincial and territorial framework for action to promote healthy weights URL: http://www.phac-aspc.gc.ca/hp-ps/hl-mvs/framework-cadre/ [accessed 2014-11-12] [WebCite Cache ID 6U2EE6UfX]
10. Viswanath K, Bond K. Social determinants and nutrition: reflections on the role of communication. J Nutr Educ Behav 2007;39(2 Suppl):S20-S24. [doi: 10.1016/j.jneb.2006.07.008] [Medline: 17336801]
11. Smith A. Older adults and technology use. Washington, DC: Pew Internet & American Life Project; 2014 Apr 03. URL: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/ [accessed 2014-04-23] [WebCite Cache ID 6P34DTel4]
12. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Internet & American Life Project; 2013 Jan 15. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [WebCite Cache ID 6P1lpw6f0]
13. Fox S. The Social Life of Health Information, 2011. Washington, DC: Pew Internet & American Life Project; 2011 May 12. URL: http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2014-04-22] [WebCite Cache ID 6P1mKBDa6]
14. Rosser BA, Vowles KE, Keogh E, Eccleston C, Mountain GA. Technologically-assisted behaviour change: a systematic review of studies of novel technologies for the management of chronic illness. J Telemed Telecare 2009;15(7):327-338. [doi: 10.1258/jtt.2009.090116] [Medline: 19815901]
15. Kerr C, Murray E, Noble L, Morris R, Bottomley C, Stevenson F, et al. The potential of Web-based interventions for heart disease self-management: a mixed methods investigation. J Med Internet Res 2010;12(4):e56 [FREE Full text] [doi: 10.2196/jmir.1438] [Medline: 21156471]
16. 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]

17. Springvloet L, Lechner L, Oenema A. Planned development and evaluation protocol of two versions of a web-based computer-tailored nutrition education intervention aimed at adults, including cognitive and environmental feedback. BMC Public Health 2014;14:47 [FREE Full text] [doi: 10.1186/1471-2458-14-47] [Medline: 24438381]
18. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
19. Chou WS, 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]
20. Williams G, Hamm MP, Shulhan J, Vandermeer B, Hartling L. Social media interventions for diet and exercise behaviours: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2014;4(2):e003926 [FREE Full text] [doi: 10.1136/bmjopen-2013-003926] [Medline: 24525388]
21. Maher CA, Lewis LK, Ferrar K, Marshall S, De BI, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
22. Anderson-Bill ES, Winett RA, Wojcik JR, Winett SG. Web-based guide to health: relationship of theoretical variables to change in physical activity, nutrition and weight at 16-months. J Med Internet Res 2011;13(1):e27 [FREE Full text] [doi: 10.2196/jmir.1614] [Medline: 21447470]
23. Neville LM, O'Hara B, Milat AJ. Computer-tailored dietary behaviour change interventions: a systematic review. Health Educ Res 2009 Aug;24(4):699-720 [FREE Full text] [doi: 10.1093/her/cyp006] [Medline: 19286893]
24. 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]
25. 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]
26. Broekhuizen K, Kroeze W, van Poppel MNM, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Ann Behav Med 2012 Oct;44(2):259-286 [FREE Full text] [doi: 10.1007/s12160-012-9384-3] [Medline: 22767052]
27. Neuenschwander LM, Abbott A, Mobley AR. Comparison of a web-based vs in-person nutrition education program for low-income adults. J Acad Nutr Diet 2013 Jan;113(1):120-126. [doi: 10.1016/j.jand.2012.07.034] [Medline: 23092741]
28. Bensley RJ, Anderson JV, Brusk JJ, Mercer N, Rivas J. Impact of internet vs traditional Special Supplemental Nutrition Program for Women, Infants, and Children nutrition education on fruit and vegetable intake. J Am Diet Assoc 2011 May;111(5):749-755. [doi: 10.1016/j.jada.2011.02.010] [Medline: 21515124]
29. Wikipedia. Blog URL: http://en.wikipedia.org/wiki/Blog [accessed 2015-03-23] [WebCite Cache ID 6XFXFWImm]
30. Lagu T, Kaufman EJ, Asch DA, Armstrong K. Content of weblogs written by health professionals. J Gen Intern Med 2008 Oct;23(10):1642-1646 [FREE Full text] [doi: 10.1007/s11606-008-0726-6] [Medline: 18649110]
31. Buis LR, Carpenter S. Health and medical blog content and its relationships with blogger credentials and blog host. Health Commun 2009 Dec;24(8):703-710. [doi: 10.1080/10410230903264014] [Medline: 20183379]
32. Denecke K, Nejdl W. How valuable is medical social media data? Content analysis of the medical web. Information Sciences 2009 May 30;179(12):1870-1880. [doi: 10.1016/j.ins.2009.01.025]
33. Miller EA, Pole A. Diagnosis blog: checking up on health blogs in the blogosphere. Am J Public Health 2010 Aug;100(8):1514-1519. [doi: 10.2105/AJPH.2009.175125] [Medline: 20558802]
34. Della LJ, Griffin DB, Eroğlu D, Bernhardt JM, Wells RR. Is there health out there in the afrosphere? An analysis of health-related content posted by black bloggers. Health Mark Q 2013;30(1):1-18. [doi: 10.1080/07359683.2013.758012] [Medline: 23458478]
35. Boepple L, Thompson JK. A content analysis of healthy living blogs: evidence of content thematically consistent with dysfunctional eating attitudes and behaviors. Int J Eat Disord 2014 May;47(4):362-367. [doi: 10.1002/eat.22244] [Medline: 24420676]
36. Gies J, Martino S. Uncovering ED: a qualitative analysis of personal blogs managed by individuals with eating disorders. The Qualitative Report 2014;19(57):1-15.
37. Lynch M. Healthy habits or damaging diets: an exploratory study of a food blogging community. Ecol Food Nutr 2010;49(4):316-335. [doi: 10.1080/03670244.2010.491054] [Medline: 21888474]
38. Cox AM, Blake MK. Information and food blogging as serious leisure. Aslib Proceedings 2011 Mar 22;63(2/3):204-220. [doi: 10.1108/00012531111135664]
39. Simunaniemi AM, Sandberg H, Andersson A, Nydahl M. Laypeople blog about fruit and vegetables for self-expression and dietary influence. Health Commun 2011 Oct;26(7):621-630. [doi: 10.1080/10410236.2011.561520] [Medline: 21541865]
40. Lynch M. From food to fuel: perceptions of exercise and food in a community of food bloggers. Health Education Journal 2010 Nov 18;71(1):72-79. [doi: 10.1177/0017896910386284]

41. Wright S, Camahort E. slideshare. 2009. 2009 Women and Social Media Study, by BlogHer, iVillage and Compass Partners URL: http://fr.slideshare.net/LisaStone/2009-women-and-social-media-study-by-blogher-ivillage-and-compass-partners [accessed 2014-04-23] [WebCite Cache ID 6P33dLFxN]
42. Lake AA, Hyland RM, Mathers JC, Rugg‐Gunn AJ, Wood CE, Adamson AJ. Food shopping and preparation among the 30‐somethings: whose job is it? (The ASH30 study). British Food Journal 2006 Jun;108(6):475-486. [doi: 10.1108/00070700610668441]
43. Flagg LA, Sen B, Kilgore M, Locher JL. The influence of gender, age, education and household size on meal preparation and food shopping responsibilities. Public Health Nutr 2014 Sep;17(9):2061-2070 [FREE Full text] [doi: 10.1017/S1368980013002267] [Medline: 23988018]
44. 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]
45. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
46. Kim M, Hunter JE. Relationships among attitudes, behavioral intentions, and behavior: a meta-analysis of past research, part 2. Communication Research 1993 Jun 01;20(3):331-364 [FREE Full text] [doi: 10.1177/009365093020003001]
47. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
48. Conner M, Norman P, Bell R. The theory of planned behavior and healthy eating. Health Psychol 2002 Mar;21(2):194-201. [Medline: 11950110]
49. Lau ASM. Hospital-based nurses' perceptions of the adoption of Web 2.0 tools for knowledge sharing, learning, social interaction and the production of collective intelligence. J Med Internet Res 2011;13(4):e92 [FREE Full text] [doi: 10.2196/jmir.1398] [Medline: 22079851]
50. Fleming ML, Barner JC, Brown CM, Shepherd MD, Strassels S, Novak S. Using the theory of planned behavior to examine pharmacists' intention to utilize a prescription drug monitoring program database. Res Social Adm Pharm 2014;10(2):285-296. [doi: 10.1016/j.sapharm.2013.05.009] [Medline: 23816495]
51. Pelling EL, White KM. The theory of planned behavior applied to young people's use of social networking Web sites. Cyberpsychol Behav 2009 Dec;12(6):755-759. [doi: 10.1089/cpb.2009.0109] [Medline: 19788377]
52. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS Quarterly 2003;27(3):425-478.
53. Cohen S. Social relationships and health. Am Psychol 2004 Nov;59(8):676-684. [doi: 10.1037/0003-066X.59.8.676] [Medline: 15554821]
54. Garriguet D. Statistics Canada. 2009. Diet quality in Canada URL: http://www.statcan.gc.ca/pub/82-003-x/2009003/article/10914-eng.pdf [accessed 2015-03-23] [WebCite Cache ID 6XFRl0zaY]
55. Patton MQ. Qualitative Research and Evaluation Methods, 3rd ed. Thousand Oaks, CA: Sage Publications; 2002.
56. Morse JM. The significance of saturation. Qualitative Health Research 1995 May 01;5(2):147-149. [doi: 10.1177/104973239500500201]
57. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
58. Godin G. Les comportements dans le domaine de la santé, first ed. Montreal, QC: Les presses de l'Université de Montréal; 2012.
59. Edwards L, Thomas C, Gregory A, Yardley L, O'Cathain A, Montgomery AA, et al. Are people with chronic diseases interested in using telehealth? A cross-sectional postal survey. J Med Internet Res 2014;16(5):e123 [FREE Full text] [doi: 10.2196/jmir.3257] [Medline: 24811914]
60. 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]
61. Greenberg S, Yaari E, Bar‐Ilan J. Perceived credibility of blogs on the internet – the influence of age on the extent of criticism. Aslib Proceedings 2013 Jan;65(1):4-18. [doi: 10.1108/00012531311297159]
62. Wetherell S. Foodista. 2013 May 29. Results of the 2013 State of Food Blogging Survey URL: http://www.foodista.com/blog/2013/05/29/results-of-the-2013-state-of-food-blogging-survey [accessed 2014-11-12] [WebCite Cache ID 6U2Ex0Zo5]
63. 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. [Medline: 23615206]
64. Zickuhr K, Madden M. Older adults and internet use. Washington, DC: Pew Internet & American Life Project; 2012 Jun 06. URL: http://www.pewinternet.org/2012/06/06/older-adults-and-internet-use/ [accessed 2014-04-23] [WebCite Cache ID 6P34T2wDK]

Abbreviations

RD: registered dietitian
TPB: Theory of Planned Behavior

Edited by G Eysenbach; submitted 30.04.14; peer-reviewed by M Lynch, B Walters; comments to author 23.09.14; revised version received 15.12.14; accepted 03.02.15; published 08.04.15.

Please cite as:
Bissonnette-Maheux V, Provencher V, Lapointe A, Dugrenier M, Dumas AA, Pluye P, Straus S, Gagnon MP, Desroches S
Exploring Women's Beliefs and Perceptions About Healthy Eating Blogs: A Qualitative Study
J Med Internet Res 2015;17(4):e87
URL: http://www.jmir.org/2015/4/e87/
doi:10.2196/jmir.3504
PMID:25858777

Original Paper

The Mediating Effect of Gaming Motivation Between Psychiatric Symptoms and Problematic Online Gaming: An Online Survey

Orsolya Király[1,2], MA; Róbert Urbán[1], PhD; Mark D Griffiths[3], PhD; Csilla Ágoston[1,2], MA; Katalin Nagygyörgy[1,2], MA; Gyöngyi Kökönyei[1], PhD; Zsolt Demetrovics[1], PhD

[1]Institute of Psychology, Eötvös Loránd University, Budapest, Hungary

[2]Doctoral School of Psychology, Eötvös Loránd University, Budapest, Hungary

[3]International Gaming Research Unit, Psychology Division, Social Sciences, Nottingham Trent University, Nottingham, United Kingdom

Corresponding Author:
Zsolt Demetrovics, PhD
Institute of Psychology
Eötvös Loránd University
1st Fl.
46 Izabella St
Budapest, 1064
Hungary
Phone: 36 30 976 1097
Fax: 36 1 461 2697
Email: demetrovics@t-online.hu

Abstract

Background: The rapid expansion of online video gaming as a leisure time activity has led to the appearance of problematic online gaming (POG). According to the literature, POG is associated with different psychiatric symptoms (eg, depression, anxiety) and with specific gaming motives (ie, escape, achievement). Based on studies of alcohol use that suggest a mediator role of drinking motives between distal influences (eg, trauma symptoms) and drinking problems, this study examined the assumption that there is an indirect link between psychiatric distress and POG via the mediation of gaming motives. Furthermore, it was also assumed that there was a moderator effect of gender and game type preference based on the important role gender plays in POG and the structural differences between different game types.

Objective: This study had two aims. The first aim was to test the mediating role of online gaming motives between psychiatric symptoms and problematic use of online games. The second aim was to test the moderator effect of gender and game type preference in this mediation model.

Methods: An online survey was conducted on a sample of online gamers (N=3186; age: mean 21.1, SD 5.9 years; male: 2859/3186, 89.74%). The Brief Symptom Inventory (BSI), the Motives for Online Gaming Questionnaire (MOGQ), and the Problematic Online Gaming Questionnaire (POGQ) were administered to assess general psychiatric distress, online gaming motives, and problematic online game use, respectively. Structural regression analyses within structural equation modeling were used to test the proposed mediation models and multigroup analyses were used to test gender and game type differences to determine possible moderating effects.

Results: The mediation models fitted the data adequately. The Global Severity Index (GSI) of the BSI indicated that the level of psychiatric distress had a significant positive direct effect (standardized effect=.35, $P<.001$) and a significant indirect (mediating) effect on POG (standardized effect=.194, $P<.001$) via 2 gaming motives: escape (standardized effect=.139, $P<.001$) and competition (standardized effect=.046, $P<.001$). The comparison of the 2 main gamer types showed no significant differences in the model. However, when comparing male and female players it was found that women had (1) slightly higher escape scores (on a 5-point Likert scale: mean 2.28, SD 1.14) than men (mean 1.87, SD 0.97) and (2) a stronger association between the escape motive and problematic online gaming (standardized effect size=.64, $P<.001$) than men (standardized effect size=.20, $P=.001$).

Conclusions: The results suggest that psychiatric distress is both directly and indirectly (via escape and competition motives) negatively associated with POG. Therefore, the exploration of psychiatric symptoms and gaming motives of POG can be helpful in the preparation of prevention and treatment programs.

(J Med Internet Res 2015;17(4):e88) doi:10.2196/jmir.3515

KEYWORDS

video games; Internet; motivation; behavior, addictive; psychopathology; coping behavior

Introduction

The large-scale expansion of the video game industry and online video gaming as a leisure time activity [1] has led to the appearance of problematic online gaming (POG) and online gaming addiction [2,3]. Although only a minority of players appear to display addiction-like symptoms and negative consequences on other important activities (ie, work, education) and relationships [4], it is still essential to explore the contributing factors in the development of this behavior. The importance of the phenomenon is also demonstrated by research findings that claim that online gaming is generally the most frequent problem associated with Internet use [5] and by the fact that POG was recently introduced into Section 3 of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*) [6] under the name of “Internet Gaming Disorder.” Hopefully its inclusion will encourage further research on the topic and help unify the field [7].

The attraction of online games appears to lie in their potential to satisfy different psychological needs that can be conceptualized as motives for gaming. In one earlier study, 7 motivational dimensions for online gaming were identified [8]. *Social motivation* concerns the need of gaming together with others and making friends. *Escape* refers to gaming in order to avoid real life problems and difficulties. *Competition* concerns the defeating of others, whereas *skill development* is about improving the player’s own coordination, concentration, and other skills. *Coping* involves stress, tension, or aggression reduction through gaming as well as getting into a better mood. *Fantasy* refers to trying out new identities and/or activities in virtual game words that are not possible in the gamers’ everyday lives. Finally, *recreation* concerns gaming for the playing of the game for fun. In fact, a recent study demonstrated that self-reported gaming motives were related to actual in-game behaviors and could also predict future in-game behaviors [9].

Studies into alcohol use have shed light on the importance of motives in both drinking and problem drinking (eg, [10]). Similarly, an association between different gaming motives and POG has been reported in the literature. Yee [11] found that escapism and advancement motives were the best predictors of problematic use of massively multiplayer online role-playing games (MMORPGs). Nagygyörgy and colleagues [12] yielded similar results (ie, escapism and achievement showed strong association with problematic MMORPG play). However, Zanetta Dauriat and colleagues [13] found that addictive MMORPG use patterns were best predicted by achievement, escapism, and socializing motives. Kwon and colleagues [14] suggested that adolescents become addicted to online games in an attempt to escape from self and reality based on Baumeister’s [15] escape theory.

It has also been reported that psychiatric symptoms can be related to POG. For instance, depression, anxiety, and social anxiety are positively related to problematic gaming [16-20]. An association has also been found between various domains of psychopathology and problem video gaming using the Symptom Checklist 90 (SCL-90) [21].

To date, psychiatric symptoms have been viewed as being directly related to POG. This approach is understandable because—as in other addictions—it is highly probable that psychiatric symptoms accompany the problem behavior. However, an indirect link might also exist between the 2 entities via the mediation of specific motivations. Such findings have been reported in the alcohol and gambling literature. Drinking motives have been shown to be the most proximate factors that precede alcohol use, the gateway via which more distal influences (eg, alcohol expectancies, anxiety sensitivity, trauma symptoms, and personality) are mediated [22-26]. For instance, the coping motive mediates between tension-reduction expectancies and drinking problems [22]. Moreover, in the case of gambling, the motivation to escape or dissociate has been found to mediate between aversive physiological and emotional states and gambling severity [27,28].

Gender differences may also play an important role in problematic gaming. Research has consistently shown that males are more likely to play online games [4,29,30] and spend significantly more time gaming than female players [31,32]. The literature also suggests that male players are at higher risk of problematic gaming than female players [4,30,32,33]. Furthermore, males and females also differ regarding their game type preference and motives to play. For instance, male gamers tend to prefer action, sports, and shooter games, whereas female gamers prefer puzzle, adventure, and quiz games [34,35]. It has also been found that male gamers score higher on motives related to achievement and competition, whereas female gamers score higher on social motives and escapism [8,11]. These findings are in-line with gender differences in competitiveness and social-emotional inclination. Males tend to be more competitive in various contexts, such as between-group interactions [36] and negotiations [37]. In contrast, females have been considered to be more interpersonally oriented [38,39]. In addition, females are more prone to experience negative emotions and to internalize problems reflected in higher prevalence of depression [40] and anxiety disorders [41], and this is in-line with their higher escapism scores. It has also been argued that clinical correlates of gaming may differ by gender [17].

Given that different game types (eg, role-playing games, first-person shooters, and real-time strategy games) vary in their basic structural characteristics and gameplay [42,43], it is reasonable to assume that psychiatric symptoms and motives related to different game types might also vary. In-line with this assumption, Stetina and colleagues [44] found that MMORPG players showed higher tendency toward depressive symptoms and escapism than gamers playing online shooters or real-time strategy games.

Consequently, the present study had 2 aims. The first aim was to test the mediating role of online gaming motives between psychiatric symptoms and problematic use of online games. Based on the literature, it was assumed that escape, coping, and competition motives would act as mediators in the model. The second aim was to test the moderator effect of gender and game type preference in this mediation model. Based on gender differences reported in the online gaming literature, differences in the paths related to competition, social, and escape motives have been expected between male and female players. On the other hand, no assumptions have been made regarding game type-related differences in the lack of findings in this domain.

Methods

The study was approved by the Institutional Review Board of the Eötvös Loránd University, Budapest, Hungary.

Recruitment

All the most frequently visited Hungarian online gaming websites having more than 100 registered users were identified (N=18). The majority of the sites (n=11) were game-specific: (massively) multiplayer online role playing games ([M]MORPGs) (eg, Lineage II, World of Warcraft, Guild Wars, Diablo II, Lord of the Rings; Thrillion kincsei [The Treasures of Thrillion])*;* (massively) multiplayer online real-time strategy ([M]MORTS) games (eg, Age of Empires 3, Travian, Klánháború [TribalWars], and Red Alert 3)*;* multiplayer online first-person shooter (MOFPS) games (eg, Call of Duty), 1 site was genre-specific (ie, role-playing games), and 6 sites were general video gaming sites. The search was carried out by 4 experts using online search engines (ie, Google) and additionally obtaining information from gamers from their circle of acquaintances. Subsequently, the 18 websites were invited to advertise for participants in the present study. All administrators agreed to cooperate and advertised the call for participation on their websites or included it in their online newsletter to gamers. In the call for participation in the study, gamers were asked to visit a password-protected website and complete the questionnaire.

Before filling out the questionnaire, all participants were informed about the general goals of the study (ie, to obtain an objective and real picture of online games and the online gamer community in contrast with the simplified and rather negative picture the media disseminates) and the time needed to complete it. They were assured about confidentiality and anonymity, and their informed consent was obtained by ticking a box if they agreed to continue and participate in the study. No personal information was collected or stored and no incentives were offered. Data collection took place between April and July 2009. The advertisements were displayed for approximately 3 months, whereas the newsletters including the call for participation were sent twice. According to the voluntary nature of participation, answering all the questions was not mandatory. The number of unique website visitors was not assessed at the time of the recruitment period, and the number of overlapping visitors between the 18 websites was also unknown; therefore, the view rate could not be calculated. In total, 7520 gamers visited the first page of the questionnaire and 4374 completed at least some of it. Consequently, the participation rate was 58.16%. Of these 4374 gamers, 3186 completed the entire survey including psychiatric symptoms, online gaming motives, and problematic online game use resulting in a completion rate of 72.84%. Consequently, all analyses were carried out on this latter subsample. Only completed questionnaires were analyzed. However, skipping answers was allowed; therefore, missing data were treated with full information maximum likelihood (FIML) method with MPlus 6.0.

Measures

Sociodemographic Variables

Data relating to major sociodemographics were collected including age, gender, marital status, and education.

Gaming-Related Variables

Data were collected regarding weekly game time and preferred game type. Game type preference was obtained using the results of a latent profile analysis on the amount of time spent playing different game types described in a previous study that used the same sample (which was slightly bigger than the current sample because it also included those respondents who did not complete the Brief Symptom Inventory [BSI] placed at the end of the questionnaire) [43]. The majority of the sample (2517/3186, 79.00%) had a clear game type preference. Almost half of the sample (1466/3186, 46.01%) played almost exclusively with MMORPGs, whereas an additional 872 of 3186 gamers (27.37%) preferred MOFPSs. A small minority of the sample, 118 of 3186 gamers (3.70%), mostly played MMORTS games. The rest of the gamers (61/3186, 1.91%) played other online games (ie, sport games, puzzle games) or did not have a clear game type preference (669/3189, 21.00%). Therefore, the majority of the sample (2338/3186, 73.38%) could be categorized as either an MMORPG player or an MOFPS player. In the moderation analysis, only these 2 groups were focused on due to the low sample size of the remaining groups (ie, MMORTS players and gamers playing other online game types).

Psychiatric Symptoms

These were assessed using the Hungarian version of the BSI [45,46], a measure that assesses self-reported clinically relevant psychological symptoms. The 53-item questionnaire uses a 5-point Likert scale (from “not at all” to “extremely”). The BSI comprises 9 symptom dimensions: somatization, obsession compulsion, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. However, in the present study, the summarized measure—the Global Severity Index (GSI)—was applied to measure the intensity of general distress. Good reliability and validity has been found across various samples [45-47] as well as in this study (Cronbach alpha=.97).

Online Gaming Motives

These were assessed using the Motives for Online Gaming Questionnaire (MOGQ) [8]. This 27-item self-report measure assesses the full range of motives for online gaming. The 7 motivational dimensions assessed by MOGQ comprise social, escape, competition, skill development, coping, fantasy, and recreation. A brief description of each motivation has already been mentioned in the Introduction. In addition, a sample item for each motivation is as follows: social (eg, “...because I can meet many different people”), escape (eg, “...to forget about unpleasant things or offenses”), competition (eg, “...because it is good to feel that I am better than others”), skill development (eg, “...because it improves my skills”), coping (eg, “...because it helps me get rid of stress”), fantasy (eg, “...to be somebody else for a while”), and recreation (eg, “...because it is entertaining”). The MOGQ uses a 5-point Likert scale from “never” to “almost always/always” with higher scores indicating higher frequency of the respective motivational dimension. Internal consistencies were reported for all 7 dimensions, ranging from .79 to .90 [8].

Problematic Use of Online Games

This was assessed using the Problematic Online Gaming Questionnaire (POGQ) [29]. The POGQ is an 18-item self-report assessment tool with good psychometric properties based on both theoretical and empirical content. It assesses 6 dimensions of problematic gaming, namely preoccupation, overuse, immersion, social isolation, interpersonal conflicts, and withdrawal. The items are measured on a 5-point Likert scale (from “never” to “always”) and a summarized score can also be calculated that shows the severity of the problematic use (with higher scores indicating more serious problems being due to online gaming). A cut-off score of 66 points has been suggested to determine the proportion of gamers at high risk of problematic use. This cut-off score was determined empirically in the same article [29]. Conducting a latent profile analysis, 4 severity groups were identified and a cut-off score was calculated by using a sensitivity and specificity analysis based on the most severe group as the criterion group. The factor structure of the POGQ has been confirmed in further studies and high internal consistencies (.91) have also been reported [4,29].

Statistical Analyses

Structural regression analyses within structural equation modeling (SEM) were used to test the proposed mediation models. Because the scales were not normally distributed, maximum likelihood estimation robust to nonnormality (MLR) was used in all SEM analyses [48]. Multigroup analyses were used to test gender and game type differences. To evaluate goodness of fit of the overall models, the chi-square goodness-of-fit statistic (with a *P* value <.05), the comparative fit index (CFI), the Tucker-Lewis fit index or nonnormed fit index (TLI or NNFI), root mean square error approximation (RMSEA) and its 90% confidence interval (90% CI), and the standardized root mean square residuals (SRMR) were used. For both CFI and TLI, values greater than 0.9 were considered a good fit, whereas both the values of RMSEA and SRMR should be less than 0.08 for an acceptable degree of fit [49,50]. Descriptive analyses were performed with the SPSS 20.0 statistical software package and all SEM analyses were performed with MPlus 6.0.

Results

Descriptive Statistics

The mean age of the sample (N=3186) was 21.1 years (SD 5.9 years, range 14-54 years), and the majority of the sample (2859/3186, 89.74%) were male. As previously mentioned, 1466 of 3186 gamers (46.01%) had a clear preference for MMORPGs and 872 of 3186 (27.37%) for MOFPSs. Among the MMORPG players, 240 gamers (16.37%) were female, whereas among MOFPS players only 20 gamers (2.29%) were female. Information regarding weekly game time and the proportion of gamers at high risk of problematic use is presented in Table 1.

It was assumed that the parameters of the proposed model might be moderated by players’ game type preference and gender. Therefore, descriptive statistics and group differences regarding the variables included in the model are presented in Tables 2 and 3.

Results demonstrated that MOFPS players reported significantly higher scores on competition, skill development, and coping motives, whereas MMORPG players scored significantly higher on fantasy, escape, POG, the GSI, and recreation.

Table 1. Weekly game time and proportion of gamers at high risk of problematic use for the overall sample, for males and females, and for multiplayer online first-person shooter (MOFPS) and massively multiplayer online role-playing games (MMORPGs) gamer types.

Gaming-related variables	Total sample, n (%) (N=3186)	Gender, n (%)		Game type preference, n (%)	
		Males (n=2859)	Females (n=327)	MOFPS players (n=872)	MMORPG players (n=1466)
Weekly game time					
<7 hours	382 (12.00)	328 (11.48)	54 (16.5)	125 (14.4)	137 (9.36)
7-14 hours	772 (24.25)	688 (24.09)	84 (25.7)	266 (30.5)	298 (20.36)
15-28 hours	1102 (34.62)	1001 (35.05)	101 (30.9)	308 (35.4)	509 (34.77)
29-42 hours	639 (20.08)	583 (20.41)	56 (17.1)	136 (15.6)	355 (24.25)
>42 hours	288 (9.05)	256 (8.96)	32 (9.8)	36 (4.1)	165 (11.27)
Gamers at high risk of problematic use[a]	2.43	2.39	2.76	1.96	2.95

[a] The proportion of gamers at high risk of problematic use was calculated using the established cut-off point (ie, 66) suggested in a previous article [29]. Note that only MOFPS and MMORPG gamers have been included because the MMORTS gamer group was very small and the remainder of the players could not be differentiated regarding their game type preference.

Table 2. Means, standard deviations (SD), and confidence intervals (CI) for multiplayer online first-person shooter (MOFPS) and massively multiplayer online role-playing games (MMORPGs) gamer types examined and for all players (MOFPS and MMORPG)[a] and effect sizes (Cohen’s *d*).

Psychopathology- and gaming-related variables[b]	All (MOFPS & MMORPG) players (n=2338)		MOFPS players (n=872)		MMORPG players (n=1466)		Comparison of MOFPS and MMORPG players		
	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	*t* (df)	*P*	*d*
Global Severity Index	0.61 (0.61)	0.59-0.64	0.57 (0.57)	0.53-0.61	0.64 (0.64)	0.61-0.67	2.82 (1986.8)	.005	0.12
POGQ Total score	36.13 (11.93)	35.65-36.62	34.99 (11.73)	34.21-35.77	36.82 (12.00)	36.20-37.44	3.59 (2326)	<.001	0.15
MOGQ									
Escape	1.93 (1.01)	1.89-1.97	1.83 (0.93)	1.76-1.88	2.00 (1.06)	1.95-2.05	4.09 (2019.5)	<.001	0.18
Coping	2.51 (1.08)	2.47-2.55	2.57 (1.12)	2.50-2.64	2.47 (1.05)	2.42-2.52	1.99 (1740.8)	.047	0.09
Fantasy	2.33 (1.13)	2.28-2.38	2.05 (1.00)	1.98-2.12	2.49 (1.17)	2.43-2.55	9.67 (2057.9)	<.001	0.40
Skill development	2.23 (1.14)	2.18-2.28	2.54 (1.21)	2.46-2.62	2.04 (1.05)	1.99-2.09	9.99 (1635.4)	<.001	0.44
Recreation	4.18 (0.87)	4.15-4.22	4.12 (0.90)	4.06-4.18	4.22 (0.85)	4.18-4.26	2.75 (1737.7)	.006	0.11
Competition	2.39 (1.19)	2.34-2.44	2.75 (1.22)	2.67-2.83	2.17 (1.11)	2.11-2.23	11.41 (1693.5)	<.001	0.50
Social	3.07 (1.20)	3.02-3.12	3.07 (1.21)	2.99-3.15	3.07 (1.19)	3.01-3.13	0.15 (2332)	.88	0.00

[a] Only MOFPS and MMORPG gamers have been included because the MMORTS gamer group was very small and the rest of the players could not be differentiated regarding their game type preference.

[b] POGQ: Problematic Online Gaming Questionnaire; MOGQ: Motives for Online Gaming Questionnaire.

Table 3. Means, standard deviations (SD), and confidence intervals (CI) for both genders and for the total sample and effects sizes (Cohen's *d*).

Psychopathology- and gaming-related variables[a]	Total sample (N=3186)		Males (n=2859)		Females (n=327)		Gender comparison		
	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	*t* (df)	*P*	*d*
Global Severity Index	0.62 (0.62)	0.60-0.64	0.60 (0.61)	0.58-0.62	0.77 (0.69)	0.70-0.85	4.05 (387.4)	<.001	0.26
POGQ Total Score	35.89 (11.85)	35.48-36.30	35.86 (11.83)	35.43-36.30	36.12 (12.31)	35.78-37.46	0.37 (3171)	.71	0.02
MOGQ									
Escape	1.91 (0.99)	1.88-1.94	1.87 (0.97)	1.84-1.90	2.28 (1.14)	2.16-2.40	6.22 (380.2)	<.001	0.39
Coping	2.50 (1.07)	2.46-2.54	2.50 (1.08)	2.46-2.54	2.56 (0.99)	2.45-2.67	1.05 (3177)	.30	0.06
Fantasy	2.33 (1.12)	2.29-2.37	2.28 (1.10)	2.24-2.32	2.77 (1.23)	2.64-2.90	6.80 (386.5)	<.001	0.42
Skill development	2.25 (1.14)	2.21-2.29	2.26 (1.15)	2.22-2.30	2.17 (1.00)	2.06-2.28	1.54 (430.9)	.13	0.08
Recreation	4.18 (0.88)	4.15-4.21	4.16 (0.88)	4.13-4.19	4.29 (0.82)	4.20-4.38	2.61 (416.5)	.009	0.15
Competition	2.41 (1.18)	2.37-2.45	2.47 (1.18)	2.43-2.51	1.80 (0.93)	1.70-1.90	12.06 (456.4)	<.001	0.63
Social	3.04 (1.20)	3.00-3.08	3.01 (1.20)	2.97-3.05	3.32 (1.18)	3.19-3.45	4.42 (3177)	<.001	0.26

[a] POGQ: Problematic Online Gaming Questionnaire; MOGQ: Motives for Online Gaming Questionnaire.

Results also demonstrated that female gamers scored significantly higher on fantasy, escape, social, and recreation motives as well as on the GSI of psychiatric symptoms, whereas male gamers reported significantly higher scores only on competition motive. The zero-order correlations between the components of the mediation model along with internal consistencies (Cronbach alpha) for all scales and subscales are presented in Table 4.

All correlation coefficients were significant at least *P*<.001 according to Bonferroni correction except for GSI-MOGQ recreation (*P*=.61).

Because both the BSI and the POGQ are multidimensional scales, a comprehensive correlation matrix is also provided as a multimedia appendix containing all subscales of the BSI, the MOGQ, and the POGQ (see Multimedia Appendix 1).

Table 4. Zero-order correlations and Cronbach alphas (N=3186).

Psychopathology- and gaming-related variables	2	3	4	5	6	7	8	9	Cronbach α
1. Global Severity Index	.55	.51	.29	.36	.11	–.01	.20	.09	.97
2. POGQ Total Score		.51	.39	.40	.19	.15	.37	.26	.91
3. MOGQ Escape			.60	.61	.24	.18	.27	.29	.87
4. MOGQ Coping				.50	.41	.41	.38	.42	.84
5. MOGQ Fantasy					.30	.29	.28	.34	.83
6. MOGQ Skill development						.23	.38	.45	.89
7. MOGQ Recreation							.23	.35	.78
8. MOGQ Competition								.32	.90
9. MOGQ Social									.91

Mediation Analysis

Overview

It was hypothesized that general psychiatric distress has both a direct and indirect effect (via the mediating effect of the 7 online gaming motives) on POG. Psychiatric distress was measured by the GSI and introduced in the model as a continuous observed variable. Problematic online gaming was measured by the summarized score of the POGQ and also introduced in the model as a continuous observed variable. Gaming motives were measured by the 7 factors of the MOGQ and introduced in the model as continuous latent variables. The proposed mediation models were tested with SEM. Because significant gender and gamer type differences were found in many of the variables included in the model, 3 different models were tested: (1) an overall model, (2) a separate multigroup analysis for the 2 gamer types, and (3) another multigroup analysis for males and females. The third model was carried out on the MMORPG gamer subsample because this group was the only one that had enough female players for comparison.

The Overall Model

The overall model had an adequate fit to the data (χ^2_{343}= 3217.8, *P*<.001; CFI=0.935; TLI=0.923; RMSEA=0.051, 90% CI 0.050-0.053; Cfit>0.90; SRMR=0.046). According to the results (see Figure 1), psychiatric symptoms had a significant direct effect on POG (standardized effect=.35, *P*<.001) as well as on all the gaming motives apart from recreation (standardized effects ranging from .10 to .55, *P*<.001). Psychiatric symptoms were significantly and strongly associated with escape, coping, and fantasy, and significantly but weakly associated with skill development, competition, and social motives. In relation to the association between gaming and problematic use, only escape and competition motives had a considerable effect size (standardized effect=.26 and .21, respectively), whereas social motives, skill development, and recreation motives had significant but low effect sizes. In relation to the indirect effect between psychiatric symptoms and problematic gaming, 4 paths were statistically significant: (1) psychiatric symptoms → escape → problematic gaming (standardized effect=.139, *P*<.001); (2) psychiatric symptoms → competition → problematic gaming (standardized effect=.046, *P*<.001); (3) psychiatric symptoms → social motives → problematic gaming (standardized effect=.008, *P*=.001); and (4) psychiatric symptoms → skill development → problematic gaming (standardized effect=-.006, *P*=.02). However, the latter 2 pathways had a negligible effect size. The mediation pathways added up a total standardized indirect effect size of .194 (*P*<.001). The proportion of the mediated effect in the total effect was 35%. Therefore, higher levels of psychiatric symptoms were associated with higher escape and competition motives that were associated with higher level of problematic use. All other indirect pathways were nonsignificant (*P*>.05) or had a negligible effect size (ie, <.01). The full model explained 44% of the total variance of POG.

Gamer Type Comparison

The model comparing MOFPS and MMORPG players also had an adequate fit to the data (χ^2_{726}=3120.7; MOFPS: χ^2_{726}=1343.6; MMORPG: χ^2_{726}=1777.0, *P*<.001; CFI=0.926; TLI=0.918; RMSEA=0.053, 90% CI 0.051-0.055; Cfit>0.90; SRMR=0.051). Overall, the results (see Figure 2) were fairly similar to the first model (Figure 1). The psychiatric symptoms had a significant direct effect on POG in both groups and the psychiatric symptoms → escape → POG and the psychiatric symptoms → competition → POG indirect pathways were significant and had a considerable effect size in both groups. In the case of MOFPS players, the standardized effect size of the direct pathway was .314 (*P*<.001), whereas the standardized effect size of the mediation pathway was .194 (*P*<.001) that amounted to 38.2% of the total effect size. The mediation pathway via escape had a standardized effect size of .119 (*P*=.001) and the one via competition had an effect size of .048 (*P*<.001). Neither of the other indirect pathways were significant. The full model explained 42% of the total variance of POG. In the case of MMORPG players, the standardized effect size of the direct pathway was .376 (*P*<.001), whereas the standardized effect size of the mediation pathway was .191 (*P*<.001) that amounted to 33.7% of the total effect size. The mediation pathway via escape had a standardized effect size of .152 (*P*<.001) and the one via competition had an effect size of .063 (*P*<.001). The full model explained 45% of the total variance of POG. Neither of the other indirect pathways was significant. The comparison of the 2 player types according to the Wald test showed no significant differences in the model.

Figure 1. The overall mediation model with standardized path coefficients and the explained variance of the endogen variables (R^2) (N=3186). All 7 mediator variables are latent variables. For clarity, indicator variables associated with them have not been depicted in this figure but were published in an earlier paper [8]. Also for clarity, the covariances between the errors of all mediator variables have not been depicted in the figure. Simple arrows: significant path coefficients, dotted arrows: nonsignificant path coefficients. *P<.05; **P<.01; ***P<.001.

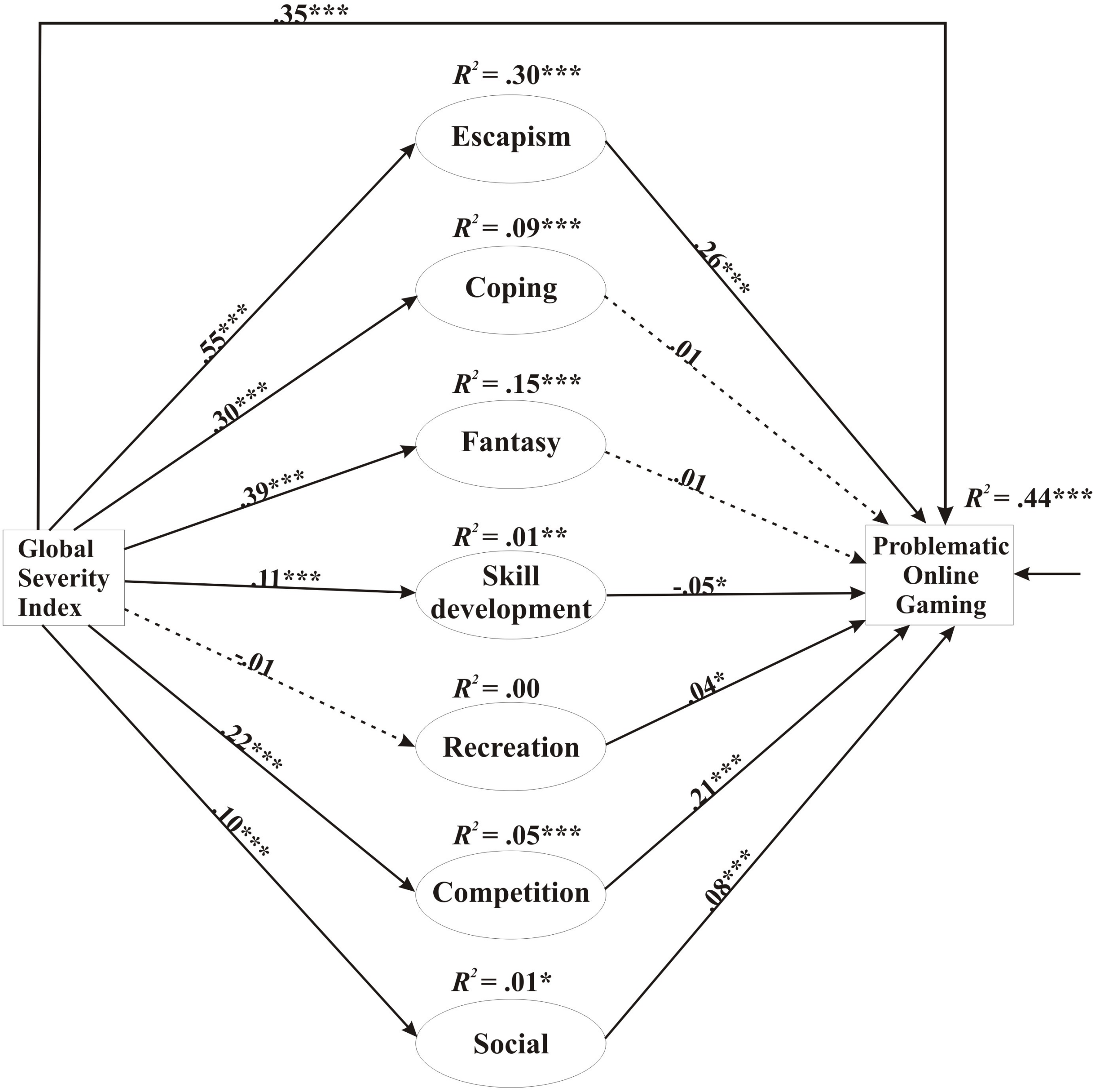
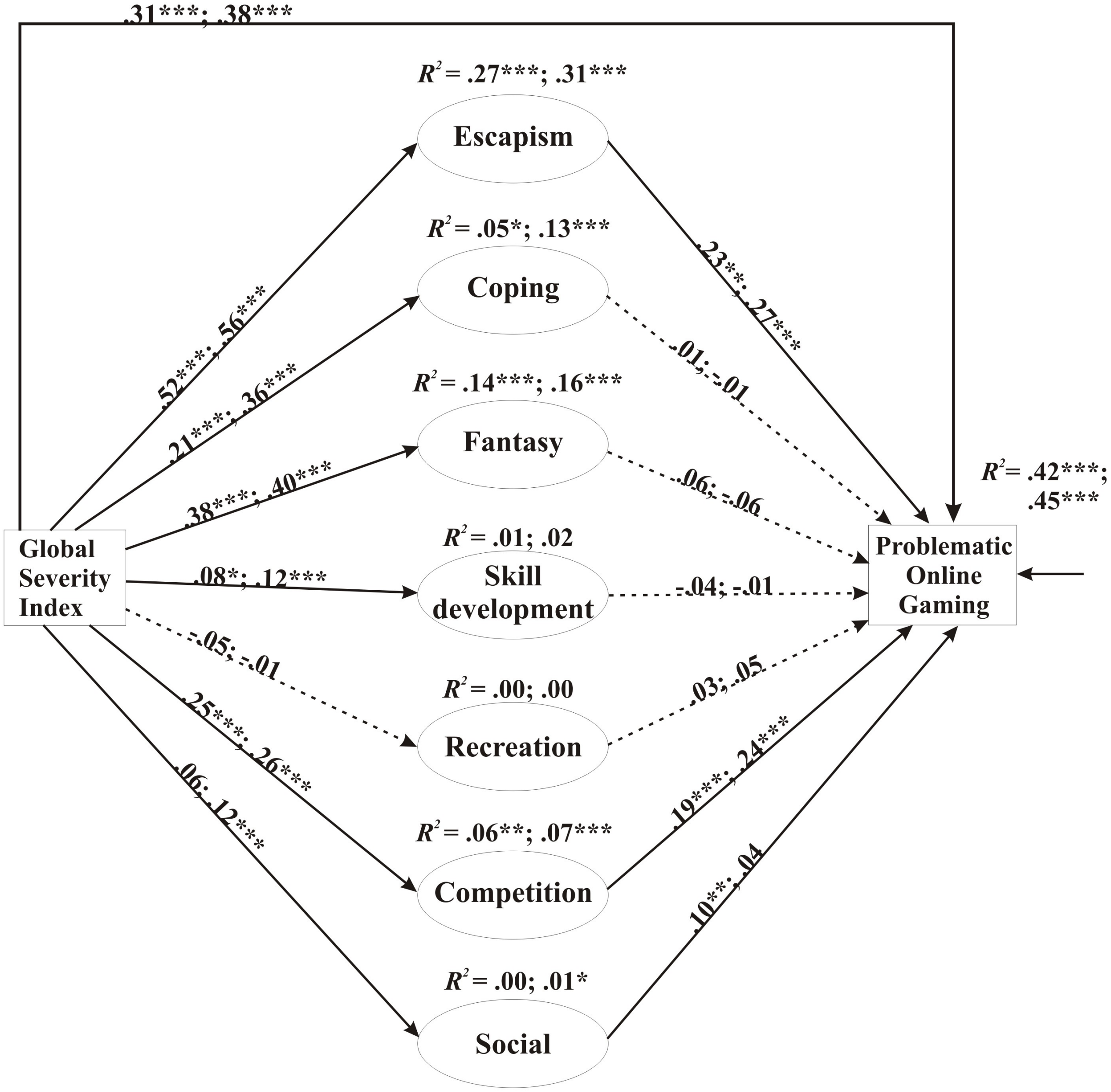


Figure 2. The mediation model and standardized path coefficients. Results of multigroup analysis and path coefficients across the 2 gamer types, multiplayer online first-person shooter (MOFPS) and massively multiplayer online role-playing games (MMORPGs), and the explained variance of the endogen variables (R^2) (MOFPS: n=872; MMORPG: n=1466). The first (left) values describe MOFPS players, whereas the second (right) values describe MMORPG players. All 7 mediator variables are latent variables. For clarity, indicator variables associated with them have not been depicted but were published in an earlier paper [8]. Also for clarity, the covariances between the errors of all mediator variables have not been depicted in the figure. Simple arrows: significant path coefficients; dotted arrows: nonsignificant path coefficients. *P<.05; **P<.01; ***P<.001.

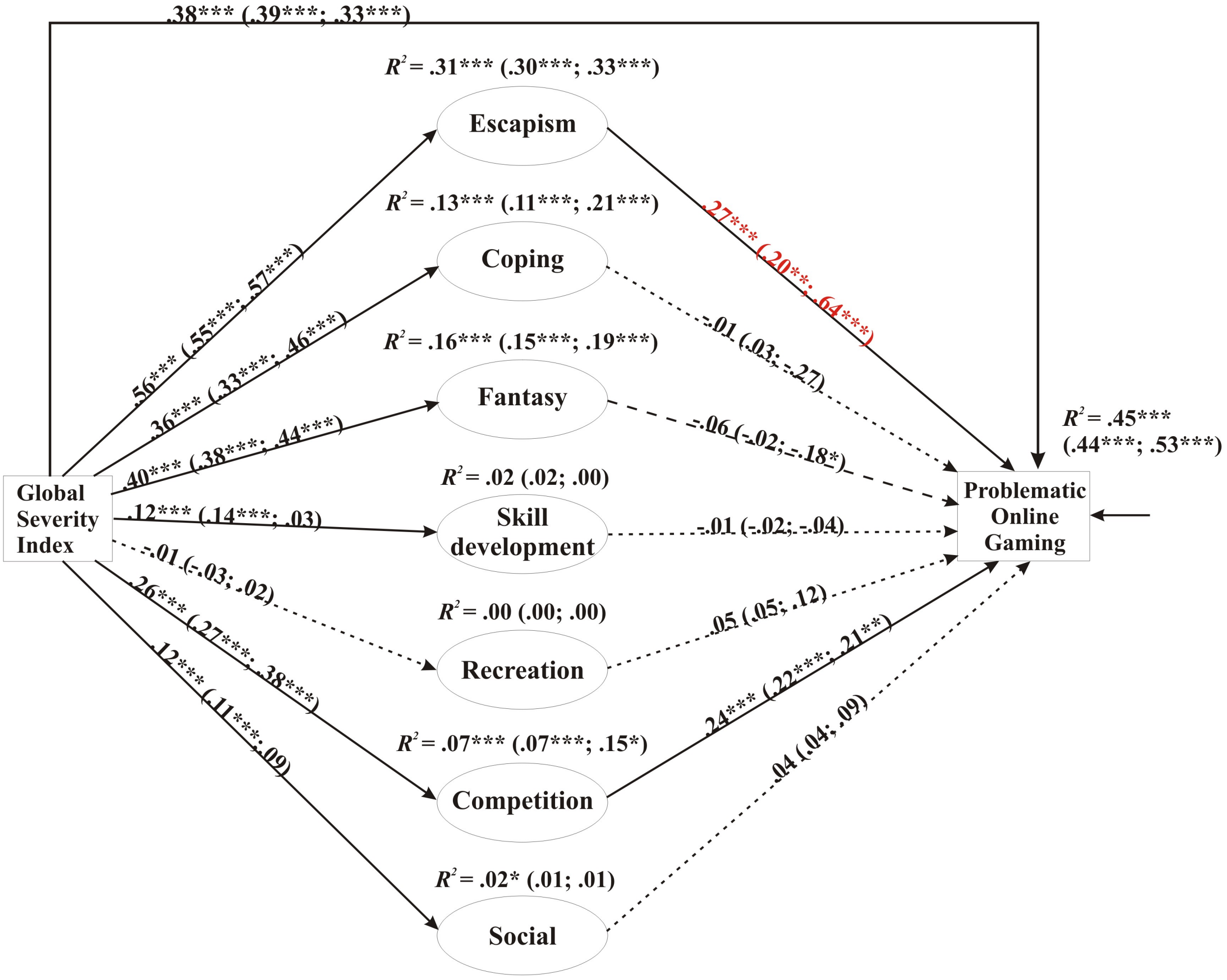


Gender Comparison Between Massively Multiplayer Online Role-Playing Game Players

Given that the proportion and sample size of female players were only considerable among MMORPG players, the third model was carried out only on this group (Figure 3). The model fitted the data properly (χ^2_{726}=2232.0; males: χ^2_{726}=1531.3; females: χ^2_{726}=700.7; P<.001; CFI=0.929; TLI=0.920; RMSEA=0.053, 90% CI 0.051-0.056; Cfit>0.90; SRMR=0.049). In relation to MMORPG players (Figure 3), the results were also similar to the results of the overall model (Figure 1). Psychiatric symptoms again had a direct effect on POG with a standardized effect size of .38 (P<.001). Again, only the mediation pathways via escape and competition were significant with standardized effect sizes of .153 (P<.001) and .063 (P<.001), respectively. The standardized effect size of the indirect link between psychiatric symptoms and POG was .191 (P<.001) that amounted to 33.7% of the total effect size. The full model explained 45% of the total variance of POG.

Results relating to gender differences showed a significant difference in the escape → POG direct link between male and female MMORPG players. The standardized direct effect size of this link was .20 (*P*=.001) for men and .64 (*P*<.001) for women. The group difference between males and females was significant (Wald test=6.11, *P*=.01). As a result, the psychiatric symptoms → escape → POG mediator pathway for female players had a much higher standardized effect size (standardized effect=.368, *P*<.001) than the one for male players (standardized effect=.111, *P*=.001). This also led to a stronger indirect link between psychiatric symptoms and problematic gaming for female players (standardized effect=.253, *P*<.001) than for male players (standardized effect=.175, *P*<.001). The total explained variance of POG by the model was also slightly higher for female players (53%) than for males (44%).

Figure 3. The mediation model and standardized path coefficients. Results of multigroup analysis and path coefficients across both genders (males/females) and the explained variance of the endogen variables (R^2) (males: n=1226; females: n=240). The first values (left of the brackets) describe all MMORPG players. The first (left) values in the brackets describe male MMORPG players, whereas the second (right) values describe female MMORPG players. The color red indicates a significant difference between male and female players according to the Wald test. All 7 mediator variables are latent variables. For clarity, indicator variables associated with them have not been depicted in this figure but were published in an earlier paper [8]. Also for clarity, the covariances between the errors of all mediator variables have not been depicted in the figure. Simple arrows: significant path coefficients; dotted arrows: nonsignificant path coefficients; dashed arrow: nonsignificant path coefficients for males, significant path coefficients for females. *P<.05; **P<.01; ***P<.001.



Discussion

Principal Results and Comparison With Prior Work

The results of the present study suggest that psychiatric symptoms are both directly and indirectly (via escape and competition motives) negatively associated with POG. The mediator effect of gaming motives amounts to approximately 35% of the total effect. To the authors' knowledge, this is the first study to statistically unravel the complex association between psychiatric symptoms, online gaming motivations, and problematic use using a SEM framework. The results relating to the associations between motives and problematic use are similar to previous findings in the psychological gaming literature. For instance, Yee [11] and Billieux and colleagues [9] found that escapism was the best predictor of problematic use in the case of MMORPG players followed by advancement motivation (ie, the desire to progress rapidly in the game—"level up"—and become powerful).

Although advancement and competition motives are not the same, they are related to each other via a common connection to achievement and performance. Progressing rapidly and gaining power eventually become a way to be competitive, a way to complete goals successfully, and a way to defeat others. However, the advancement motive was developed in studies focusing on MMORPGs where "leveling up" is of a particular importance due to the persistence of the virtual world. On the other hand, the competition motive used in the present study is related to online games in general including games such as multiplayer online first-person shooter games or strategy games, in addition to MMORPGs. Other studies [12,13] have also reported a strong association between escapism and problematic gaming as well as achievement and problematic use. Achievement is the higher order factor in Yee's [11] motivational model comprising 3 subdimensions (ie, advancement, mechanics, and competition), whereas in the study by Zanetta Dauriat [13], the motive with the same name refers to the need for being competitive, to obtain fame and recognition, and to be member of a top guild. Furthermore, our results relating to the association between psychiatric symptoms and motives are also in-line with previous findings. For instance, Hagström and Kaldo [51] reported that among all online gaming motives, escapism showed the strongest relationship with psychological distress.

In addition to confirming previous findings in the gaming literature, as predicted the present study showed that the same 2 motives (escape and competition) mediated between psychiatric symptoms and problematic gaming. Playing games to escape everyday difficulties appears to be a motivating behavior that can ease psychiatric distress, and thus extends self-medication theory [52] to online gaming. This theory states that substance use is a coping strategy through which users try to compensate their psychiatric distress and attain emotional stability. This compensatory behavior then contributes to the development and maintenance of the problem behavior. The findings outlined in the present study also strengthen the inclusion of escapism as an individual criterion for Internet Gaming Disorder in *DSM-5* [6].

The second mediator variable in the present study was competition. Despite the fact that competition is usually considered a healthy and adaptive behavior, our findings suggest that in some cases it can also be a pathological factor. This has also been reported in the literature on problem gambling in which problem gamblers have been shown to be more competitive than nonproblem gamblers [53]. Gamers whose psychiatric distress level is high might use online gaming as a source for achievement through defeating other players and winning in general. If games are the only sources that maintain and boost their self-confidence and self-efficacy, and thus become a replacement for real life competition and achievement, the activity appears to increase the risk of developing a problematic behavior. However, this reasoning needs further confirmation.

In contrast to prestudy expectations, coping did not mediate between psychiatric distress and problematic gaming. Earlier motivational research yielded the surprising finding that although highly correlated (.60), coping and escape are distinct motives [8], and the present study strengthened the argument that these 2 motives have different mechanisms of action. The results of the present study suggest that in contrast to playing to escape everyday problems, gaming can also be used as an adaptive coping strategy for stress release or tension reduction without necessarily leading to problematic use. A possible explanation might be that different underlying mechanisms lie behind the 2 strategies. Avoiding real life problems (ie, escape) only alleviates the perceived stress for a short time, retaining or further multiplying the original problem (ie, stress source). On the other hand, channeling everyday stress, tension, or aggression into gaming (ie, coping) can be an active coping mechanism where at least some extent of the perceived stress is dissipated while playing. However, this is speculative and requires further research.

The recreational use of online games was related neither to psychiatric symptoms nor to problematic use of games. This suggests that playing online games can be a healthy form of entertainment if it is used moderately and balanced with other leisure time activities (ie, sports). This result also serves as counterweight for media scaremongering that often exaggerates the potential dangers of video games [54].

This model in the present study was found to be invariant across game type preference (ie, MOFPS or MMORPG), but varied significantly between males and females in the case of MMORPG players. As expected, females were characterized by a stronger link between escape and POG, and also had higher escape scores than males. This latter result is in-line with Yee's findings [55] that examined the motivational background of MMORPG players and also found that female players scored higher on the escape motive than male players. However, the present study suggests that this higher inclination for escape motivation among females is linked to a higher risk of problematic use. In contrast to the other prestudy assumptions, no gender differences were found regarding competition and social motives.

In addition, it is important to point out that the proportion of female players in the present sample is much lower than the proportion of women that play video games in the general population (ie, approximately 40% [4,18]). This is most likely due to the online data collection method in which participation is voluntary. In online gamer samples, the proportion of female players is usually quite low (approximately 10%-20% [55,56]). The reason might be that the so-called "hard-core" gamers are more interested in participating in such research studies and that the proportion of hard-core female players is lower than the proportion of hard-core male players. Therefore, it is important to acknowledge that these findings apply more to those female gamers who play seriously (in a hard-core or "masculine" gamer manner) than to the average female (casual) gamers in the general population [57,58].

Limitations

Despite the advantage of a large sample size, the self-selected and self-reported nature of the Hungarian-only data need to be taken into consideration when generalizing the results (especially because recent research has shown that self-selection of MMORPG players affects the sample's representativeness [59]). Consequently, there is a clear need for future observational and clinical studies to confirm the findings of the present study in other nationalities of gamers. The cross-sectional study design should be also borne in mind when applying the findings because causation and directionality of findings cannot be confirmed. Consequently, future studies should also use longitudinal or experimental design to establish causal relations regarding the proposed model. Furthermore, it is theoretically feasible to construct alternative models that act in the opposite direction (ie, POG leading to psychiatric distress) and/or that both models may not be mutually exclusive (that for some people psychiatric distress leads to POG and vice versa for others). Such possibilities should also be empirically tested in future studies.

Implications for Prevention and Treatment

The present study has some direct implications for prevention and treatment. There is little reason for parents, educators, and health professionals to be concerned or worried about the recreational use of online games. Neither should they necessarily be concerned about playing as a way to cope with day-to-day stress or tension. However, playing excessively as a way to avoid real-life problems or to defeat other players should receive attention because such motivations may lead to negative (addiction-like) real-life consequences. Therefore, exploring gaming motivations both on the individual and group level are likely to be helpful in the preparation of prevention and treatment programs concerning problematic gaming.

Acknowledgments

Present work was supported by the Hungarian Scientific Research Fund (grant numbers: 83884 and 111938). Zsolt Demetrovics and Gyöngyi Kökönyei acknowledge financial support of the János Bolyai Research Fellowship awarded by the Hungarian Academy of Science.

Authors' Contributions

KN, KG, and ZD designed the study; CA was responsible for data collection; OK and RU run the analysis; and OK, MDG, and ZD drafted the manuscript. All authors commented on the draft and contributed to the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Correlation matrix including the nine Brief Symptom Inventory (BSI) subscales, the seven Motives for Online Gaming Questionnaire (MOGQ) subscales and the six Problematic Online Gaming Questionnaire (POGQ) subscales.

[XLS File (Microsoft Excel File), 47KB - jmir_v17i4e88_app1.xls]

References

1. Gaudiosi J. Forbes. 2012. New reports forecast global video game industry will reach $82 billion by 2017 URL: http://www.forbes.com/sites/johngaudiosi/2012/07/18/new-reports-forecasts-global-video-game-industry-will-reach-82-billion-by-2017/ [accessed 2014-04-10] [WebCite Cache ID 6OjUvhrdh]
2. Kuss DJ, Griffiths MD. Internet gaming addiction: A systematic review of empirical research. Int J Ment Health Addiction 2011 Mar 16;10(2):278-296. [doi: 10.1007/s11469-011-9318-5]
3. Kuss DJ, Griffiths MD. Online gaming addiction in children and adolescents: A review of empirical research. Journal of Behavioral Addictions 2012 Mar 29;1(1):3-22. [doi: 10.1556/JBA.1.2012.1.1]
4. Pápay O, Urbán R, Griffiths MD, Nagygyörgy K, Farkas J, Kökönyei G, et al. Psychometric properties of the problematic online gaming questionnaire short-form and prevalence of problematic online gaming in a national sample of adolescents. Cyberpsychol Behav Soc Netw 2013 May;16(5):340-348 [FREE Full text] [doi: 10.1089/cyber.2012.0484] [Medline: 23621688]
5. Thorens G, Achab S, Billieux J, Khazaal Y, Khan R, Pivin E, et al. Characteristics and treatment response of self-identified problematic Internet users in a behavioral addiction outpatient clinic. J Behav Addict 2014 Mar;3(1):78-81 [FREE Full text] [doi: 10.1556/JBA.3.2014.008] [Medline: 25215217]
6. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Washington, DC: American Psychiatric Association; 2013.
7. Griffiths MD, King DL, Demetrovics Z. DSM-5 internet gaming disorder needs a unified approach to assessment. Neuropsychiatry 2014 Feb;4(1):1-4. [doi: 10.2217/npy.13.82]

8. Demetrovics Z, Urbán R, Nagygyörgy K, Farkas J, Zilahy D, Mervó B, et al. Why do you play? The development of the Motives for Online Gaming Questionnaire (MOGQ). Behav Res Methods 2011 Sep;43(3):814-825. [doi: 10.3758/s13428-011-0091-y] [Medline: 21487899]
9. Billieux J, Van der Linden M, Achab S, Khazaal Y, Paraskevopoulos L, Zullino D, et al. Why do you play World of Warcraft? An in-depth exploration of self-reported motivations to play online and in-game behaviours in the virtual world of Azeroth. Comput Hum Behav 2013 Jan;29(1):103-109. [doi: 10.1016/j.chb.2012.07.021]
10. Kuntsche E, Knibbe R, Gmel G, Engels R. Why do young people drink? A review of drinking motives. Clin Psychol Rev 2005 Nov;25(7):841-861. [doi: 10.1016/j.cpr.2005.06.002] [Medline: 16095785]
11. Yee N. Motivations for play in online games. Cyberpsychol Behav 2006 Dec;9(6):772-775. [doi: 10.1089/cpb.2006.9.772] [Medline: 17201605]
12. Nagygyörgy K, Mihalik Á, Demetrovics Z. Az online játékok pszichológiai vonatkozásai [Psychological aspects of online games]. In: Gabos E, editor. A média hatása a gyermekekre és fiatalokra VI. Budapest: Nemzetközi Gyermekmentő Szolgálat Magyar Egyesülete; 2012:242-248.
13. Zanetta DF, Zermatten A, Billieux J, Thorens G, Bondolfi G, Zullino D, et al. Motivations to play specifically predict excessive involvement in massively multiplayer online role-playing games: evidence from an online survey. Eur Addict Res 2011;17(4):185-189. [doi: 10.1159/000326070] [Medline: 21494046]
14. Kwon JH, Chung CS, Lee J. The effects of escape from self and interpersonal relationship on the pathological use of Internet games. Community Ment Health J 2011 Feb;47(1):113-121. [doi: 10.1007/s10597-009-9236-1] [Medline: 19701792]
15. Baumeister RF. Suicide as escape from self. Psychol Rev 1990 Jan;97(1):90-113. [Medline: 2408091]
16. Brunborg GS, Mentzoni RA, Frøyland LR. Is video gaming, or video game addiction, associated with depression, academic achievement, heavy episodic drinking, or conduct problems? J Behav Addict 2014 Mar;3(1):27-32 [FREE Full text] [doi: 10.1556/JBA.3.2014.002] [Medline: 25215212]
17. Desai RA, Krishnan-Sarin S, Cavallo D, Potenza MN. Video-gaming among high school students: health correlates, gender differences, and problematic gaming. Pediatrics 2010 Dec;126(6):e1414-e1424 [FREE Full text] [doi: 10.1542/peds.2009-2706] [Medline: 21078729]
18. Mentzoni RA, Brunborg GS, Molde H, Myrseth H, Skouverøe KJM, Hetland J, et al. Problematic video game use: estimated prevalence and associations with mental and physical health. Cyberpsychol Behav Soc Netw 2011 Oct;14(10):591-596. [doi: 10.1089/cyber.2010.0260] [Medline: 21342010]
19. Peng W, Liu M. Online gaming dependency: a preliminary study in China. Cyberpsychol Behav Soc Netw 2010 Jun;13(3):329-333. [Medline: 20557254]
20. Walther B, Morgenstern M, Hanewinkel R. Co-occurrence of addictive behaviours: personality factors related to substance use, gambling and computer gaming. Eur Addict Res 2012;18(4):167-174 [FREE Full text] [doi: 10.1159/000335662] [Medline: 22398819]
21. Starcevic V, Berle D, Porter G, Fenech P. Problem video game use and dimensions of psychopathology. Int J Ment Health Addiction 2010 Jul 7;9(3):248-256. [doi: 10.1007/s11469-010-9282-5]
22. Cooper ML, Frone MR, Russell M, Mudar P. Drinking to regulate positive and negative emotions: a motivational model of alcohol use. J Pers Soc Psychol 1995 Nov;69(5):990-1005. [Medline: 7473043]
23. Kaysen D, Dillworth TM, Simpson T, Waldrop A, Larimer ME, Resick PA. Domestic violence and alcohol use: trauma-related symptoms and motives for drinking. Addict Behav 2007 Jun;32(6):1272-1283 [FREE Full text] [doi: 10.1016/j.addbeh.2006.09.007] [Medline: 17098370]
24. Kuntsche E, Knibbe R, Engels R, Gmel G. Drinking motives as mediators of the link between alcohol expectancies and alcohol use among adolescents. J Stud Alcohol Drugs 2007 Jan;68(1):76-85. [Medline: 17149520]
25. Stewart SH, Zvolensky MJ, Eifert GH. Negative-reinforcement drinking motives mediate the relation between anxiety sensitivity and increased drinking behavior. Personality and Individual Differences 2001 Jul;31(2):157-171. [doi: 10.1016/S0191-8869(00)00213-0]
26. Urbán R, Kökönyei G, Demetrovics Z. Alcohol outcome expectancies and drinking motives mediate the association between sensation seeking and alcohol use among adolescents. Addict Behav 2008 Oct;33(10):1344-1352. [doi: 10.1016/j.addbeh.2008.06.006] [Medline: 18619739]
27. Gupta R, Derevensky JL. An empirical examinination of Jacobs' General Theory of Addictions: Do adolescent gamblers fit the theory? J Gambl Stud 1998;14(1):17-49. [Medline: 12766433]
28. Jacobs DF. A general theory of addictions: A new theoretical model. J Gambling Stud 1986;2(1):15-31. [doi: 10.1007/BF01019931]
29. Demetrovics Z, Urbán R, Nagygyörgy K, Farkas J, Griffiths MD, Pápay O, et al. The development of the Problematic Online Gaming Questionnaire (POGQ). PLoS One 2012;7(5):e36417 [FREE Full text] [doi: 10.1371/journal.pone.0036417] [Medline: 22590541]
30. Gentile DA, Choo H, Liau A, Sim T, Li D, Fung D, et al. Pathological video game use among youths: a two-year longitudinal study. Pediatrics 2011 Feb;127(2):e319-e329 [FREE Full text] [doi: 10.1542/peds.2010-1353] [Medline: 21242221]
31. Gentile D. Pathological video-game use among youth ages 8 to 18: a national study. Psychol Sci 2009 May;20(5):594-602. [doi: 10.1111/j.1467-9280.2009.02340.x] [Medline: 19476590]

32. Rehbein F, Kleimann M, Mössle T. Prevalence and risk factors of video game dependency in adolescence: results of a German nationwide survey. Cyberpsychol Behav Soc Netw 2010 Jun;13(3):269-277. [Medline: 20557246]
33. Lemmens JS, Valkenburg PM, Peter J. Psychosocial causes and consequences of pathological gaming. Comp Hum Behav 2011;27(1):144-152. [doi: 10.1016/j.chb.2010.07.015]
34. Chou C, Tsai M. Gender differences in Taiwan high school students’ computer game playing. Comp Hum Behav 2007 Jan;23(1):812-824. [doi: 10.1016/j.chb.2004.11.011]
35. Karakus T, Inal Y, Cagiltay K. A descriptive study of Turkish high school students’ game-playing characteristics and their considerations concerning the effects of games. Comp Hum Behav 2008 Sep;24(6):2520-2529. [doi: 10.1016/j.chb.2008.03.011]
36. Pemberton MB, Insko CA, Schopler J. Memory for and experience of differential competitive behavior of individuals and groups. J Pers Soc Psychol 1996 Nov;71(5):953-966. [Medline: 8939043]
37. Walters AE, Stuhlmacher AF, Meyer LL. Gender and negotiator competitiveness: a meta-analysis. Organ Behav Hum Decis Process 1998 Oct;76(1):1-29. [Medline: 9756737]
38. Baumeister RF, Sommer KL. What do men want? Gender differences and two spheres of belongingness: comment on Cross and Madson (1997). Psychol Bull 1997 Jul;122(1):38-44; discussion 51. [Medline: 9204778]
39. Taylor SE, Klein LC, Lewis BP, Gruenewald TL, Gurung RA, Updegraff JA. Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight. Psychol Rev 2000 Jul;107(3):411-429. [Medline: 10941275]
40. Grigoriadis S, Robinson GE. Gender issues in depression. Ann Clin Psychiatry 2007 Dec;19(4):247-255. [doi: 10.1080/10401230701653294] [Medline: 18058282]
41. Pigott TA. Gender differences in the epidemiology and treatment of anxiety disorders. J Clin Psychiatry 1999;60 Suppl 18:4-15. [Medline: 10487250]
42. King D, Delfabbro P, Griffiths M. Video game structural characteristics: A new psychological taxonomy. Int J Ment Health Addiction 2009 Apr 7;8(1):90-106. [doi: 10.1007/s11469-009-9206-4]
43. Nagygyörgy K, Urbán R, Farkas J, Griffiths MD, Zilahy D, Kökönyei G, et al. Typology and sociodemographic characteristics of massively multiplayer online game players. International Journal of Human-Computer Interaction 2013 Jan;29(3):192-200. [doi: 10.1080/10447318.2012.702636]
44. Stetina BU, Kothgassner OD, Lehenbauer M, Kryspin-Exner I. Beyond the fascination of online-games: Probing addictive behavior and depression in the world of online-gaming. Comp Hum Behav 2011 Jan;27(1):473-479. [doi: 10.1016/j.chb.2010.09.015]
45. Derogatis LR. Brief Symptom Inventory. Baltimore, MD: Clinical Psychometric Research; 1975.
46. Urbán R, Kun B, Farkas J, Paksi B, Kökönyei G, Unoka Z, et al. Bifactor structural model of symptom checklists: SCL-90-R and Brief Symptom Inventory (BSI) in a non-clinical community sample. Psychiatry Res 2014 Apr 30;216(1):146-154. [doi: 10.1016/j.psychres.2014.01.027] [Medline: 24524946]
47. Derogatis LR. BSI Brief Symptom Inventory. Administration, Scoring, and Procedures Manual. 4th edition. Minneapolis, MN: National Computer Systems; 1993.
48. Muthén LK, Muthén BO. Mplus User’s Guide. Fifth Edition. Los Angeles, CA: Muthén & Muthén; 1998-2007.
49. Browne M, Cudek R. Alternative ways of assessing model fit. In: Bollen K, Long J, editors. Testing Structural Equation Models. Newbury Park, CA: Sage; 1993:136-162.
50. Kline RB. Principles and Practice of Structural Equation Modeling. 2nd edition. New York: Guilford Press; 2005.
51. Hagström D, Kaldo V. Escapism among players of MMORPGs--conceptual clarification, its relation to mental health factors, and development of a new measure. Cyberpsychol Behav Soc Netw 2014 Jan;17(1):19-25. [doi: 10.1089/cyber.2012.0222] [Medline: 24003967]
52. Khantzian EJ. The self-medication hypothesis of addictive disorders: focus on heroin and cocaine dependence. Am J Psychiatry 1985 Nov;142(11):1259-1264. [Medline: 3904487]
53. Parke A, Griffiths M, Irwing P. Personality traits in pathological gambling: Sensation seeking, deferment of gratification and competitiveness as risk factors. Addict Res Theory 2004 Jan;12(3):201-212. [doi: 10.1080/1606635310001634500]
54. Griffiths MD. The role of context in online gaming excess and addiction: Some case study evidence. Int J Ment Health Addiction 2009 Jul 7;8(1):119-125. [doi: 10.1007/s11469-009-9229-x]
55. Yee N. The demographics, motivations and derived experiences of users of massively-multiuser online graphical environments. Presence-Teleop Virt 2006;15(3):309-329.
56. Griffiths MD, Davies MNO, Chappell D. Breaking the stereotype: the case of online gaming. Cyberpsychol Behav 2003 Feb;6(1):81-91. [doi: 10.1089/109493103321167992] [Medline: 12650566]
57. Griffiths MD, Lewis A. Confronting gender representation: A qualitative study of the experiences and motivations of female casual-gamers. Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport 2011;28:245-272.
58. McLean L, Griffiths MD. Female gamers: A thematic analysis of their gaming experience. International Journal of Game-Based Learning 2013;3(3):54-71.
59. Khazaal Y, van Singer M, Chatton A, Achab S, Zullino D, Rothen S, et al. Does self-selection affect samples' representativeness in online surveys? An investigation in online video game research. J Med Internet Res 2014;16(7):e164 [FREE Full text] [doi: 10.2196/jmir.2759] [Medline: 25001007]

Abbreviations

BSI: Brief Symptom Inventory
CFI: comparative fit index
FIML: full information maximum likelihood
GSI: Global Severity Index
MMORPG: massively multiplayer online role-playing games
MOFPS: multiplayer online first-person shooter
MMORTS: massively multiplayer online real-time strategy
MOGQ: Motives for Online Gaming Questionnaire
NNFI: nonnormed fit index
POG: problematic online gaming
POGQ: Problematic Online Gaming Questionnaire
RMSEA: root mean square error approximation
SEM: structural equation modeling
SRMR: standardized root mean square residuals
TLI: Tucker-Lewis fit index

Edited by G Eysenbach; submitted 07.05.14; peer-reviewed by J Billieux, R Mentzoni; comments to author 19.09.14; revised version received 09.11.14; accepted 20.12.14; published 07.04.15.

Please cite as:
Király O, Urbán R, Griffiths MD, Ágoston C, Nagygyörgy K, Kökönyei G, Demetrovics Z
The Mediating Effect of Gaming Motivation Between Psychiatric Symptoms and Problematic Online Gaming: An Online Survey
J Med Internet Res 2015;17(4):e88
URL: http://www.jmir.org/2015/4/e88/
doi:10.2196/jmir.3515
PMID:25855558

Original Paper

The Painful Tweet: Text, Sentiment, and Community Structure Analyses of Tweets Pertaining to Pain

Patrick J Tighe[1], MD, MS; Ryan C Goldsmith[1], MD; Michael Gravenstein[1], MD; H Russell Bernard[2], PhD; Roger B Fillingim[3], PhD

[1]University of Florida College of Medicine, Department of Anesthesiology, Gainesville, FL, United States
[2]University of Florida, Department of Anthropology, Gainesville, FL, United States
[3]University of Florida, Department of Community Dentistry and Behavioral Science, Gainesville, FL, United States

Corresponding Author:
Patrick J Tighe, MD, MS
University of Florida College of Medicine
Department of Anesthesiology
1600 SW Archer Road
PO Box 100254
Gainesville, FL, 32610
United States
Phone: 1 352 273 7844
Fax: 1 352 392 7029
Email: ptighe@anest.ufl.edu

Abstract

Background: Despite the widespread popularity of social media, little is known about the extent or context of pain-related posts by users of those media.

Objective: The aim was to examine the type, context, and dissemination of pain-related tweets.

Methods: We used content analysis of pain-related tweets from 50 cities to unobtrusively explore the meanings and patterns of communications about pain. Content was examined by location and time of day, as well as within the context of online social networks.

Results: The most common terms published in conjunction with the term “pain” included feel (n=1504), don’t (n=702), and love (n=649). The proportion of tweets with positive sentiment ranged from 13% in Manila to 56% in Los Angeles, CA, with a median of 29% across cities. Temporally, the proportion of tweets with positive sentiment ranged from 24% at 1600 to 38% at 2400, with a median of 32%. The Twitter-based social networks pertaining to pain exhibited greater sparsity and lower connectedness than did those social networks pertaining to common terms such as apple, Manchester United, and Obama. The number of word clusters in proportion to node count was greater for emotion terms such as tired (0.45), happy (0.43), and sad (0.4) when compared with objective terms such as apple (0.26), Manchester United (0.14), and Obama (0.25).

Conclusions: Taken together, our results suggest that pain-related tweets carry special characteristics reflecting unique content and their communication among tweeters. Further work will explore how geopolitical events and seasonal changes affect tweeters’ perceptions of pain and how such perceptions may affect therapies for pain.

(J Med Internet Res 2015;17(4):e84) doi:10.2196/jmir.3769

KEYWORDS

Twitter messaging; emotions; text mining; social networks

Introduction

Twitter is the most popular microblogging website in the world, with more than 1 billion tweets posted every 3 days [1]. The Twitter application programming interface (API) permits researchers to search for keywords in content that is posted in short, 140-character “tweets” written from a variety of fixed locations and mobile computing platforms, thus offering insights into the day-to-day discourse of personal and geopolitical events [2-5]. This API also captures information pertaining to “retweets” and “mentions,” whereby a Twitter user specifically tags a tweet to another individual, which permits the tracking of Twitter communities. These unique characteristics of Twitter have spurred investigations of tweets on topics ranging from politics, finance, and sports to health-related issues, such as H1N1 influenza epidemiology, tobacco cessation, and disaster response [2,3,6-12]. To the best of our knowledge, there have been no investigations into how pain has been discussed across Twitter.

For many, pain represents a ubiquitous yet thankfully transient experience in everyday life. However, for more than 100 million Americans, an otherwise transient experience fails to subside, progressing into chronic pain conditions costing upwards of US $635 billion dollars. More than half of hospitalized patients and 50%-75% of cancer patients die while suffering from moderate to severe pain. In the acute pain setting, more than 60% of surgical patients suffer from moderate to severe pain following surgery [14,15]. Depending on the type of surgery, up to 50% of patients will progress directly to a chronic pain condition stemming from their surgery [14,15]. A wealth of evidence points to many specific psychosocial factors that modulate pain intensity. Given the strong emotive content of social media postings, it seems prudent to investigate how pain is discussed across widespread social media platforms such as Twitter.

Two core domains of Twitter content exploration are (1) content analysis or the extraction of meaning from the tweet itself and (2) community structure analysis or the measurement of social networks based on patterns of retweets among tweeters. Content analysis includes simple measurement of word use and association along with quantification of tweet affect via sentiment analysis [16-21]. Using rules and statistical modeling techniques developed on manually annotated corpora, or body, of texts, tweets can be classified as positive in sentiment (eg, “No pain no gain, great workout, I love exercise! ”) or negative in sentiment (eg, “Ouch, my back really hurts, so sad I will miss soccer practice, bummer! ”) [4,16,17,22-29]. Content analysis can offer insight into how tweeters incorporate the term “pain” into their daily tweets, measuring the concepts discussed and the emotional tags implicit within such tweets.

Community structure analyses of retweets measure the connectedness of Twitter-based social networks. Prior observations suggest that chronic pain may be associated with, or even induce, different forms of social isolation [30-36]. Contrary to that, pain itself may be a focus of commiseration as a pointed topic of discussion, such that psychosocial therapies often focus on improving social support systems and/or minimizing social isolation [37,38]. Community structure analyses of retweet patterns for pain-related tweets can help determine whether online communications about pain reflect a more limited, intimate network of communiqués versus a more expansive dissemination of pain-related content.

To the best of our knowledge, such analytic approaches have yet to be applied to tweeters’ communiqués pertaining to pain. Here, we explore the content of more than 65,000 tweets from around the world, each containing the term “pain.” We discuss the possibility that using a combination of text analysis and network analysis, Twitter can be leveraged to study the qualitative, multidimensional facets of pain unobtrusively in the context of daily living. We had 2 objectives: (1) to evaluate the context and sentiment of pain-related tweets and (2) to compare the connectedness of Twitter-based social networks pertaining to pain against those networks containing cross-culturally universal emotion terms (happy, excitement, sad, fear, tired, anguish) and a sample of common objective terms (apple, Manchester United, Obama) [39,40]. Hypotheses pertaining to these objectives were as follows:

1. We hypothesized the existence of specific topics associated with pain; in the null, a random set of terms associated with pain.
2. We hypothesized a mixture of positive and negative sentiment in pain-related tweets; in the null, a uniformly negative sentiment.
3. We hypothesized a unique connectivity pattern of Twitter-based retweet networks pertaining to pain when compared with networks pertaining to other emotive and nonemotive terms; in the null, a connectivity pattern indistinguishable from other Twitter-based retweet networks.

Methods

Overview

The institutional review board at the University of Florida (IRB-02) declared this project exempt as a survey study on public data. Two series of analyses were conducted. The first examined the content of tweets related to pain and the second explored the social networks of tweeters retweeting content related to pain. Each set of analyses employed a separate corpus of pain-related tweets.

Content Analysis of Tweets

Overview

In classical content analysis, human readers identify the themes or concepts in a set of texts. We used automated, computer-based content analysis to extract the concepts mentioned frequently in 65,000 tweets pertaining to pain. This approach to extracting concepts from social media platforms has previously been demonstrated by a number of teams addressing a wide range of issues [12,23,26,41-46]. We also analyzed the context of pain-related words to distinguish between positive and negative uses of those words. Here, we describe our approach to the content analysis of pain-related tweets with a special emphasis on the quantification of the sentiment used within these pain-related tweets. Additional technical details are available in Multimedia Appendix 1.

Pain Tweet Corpus Generation

Data were collected during a single search in September of 2012. We first created a function to request 1500 of the most recent English-language tweets containing the term “pain” [9]. The date and time of posting of each tweet were collected. The time of day documented for each tweet was then adjusted to reflect the local hour of time for the city in question when the tweet was posted. To identify the city of origin for each tweet, the function searched for tweets posted from within a 100-mile radius of the latitude and longitude specified for a set of 50 large, English-speaking cities from around the world. Although this approach provides a geolocation for each tweet, it is important to recognize that this approach did not capture nongeolocated tweets, potentially biasing the results toward those individuals with more sophisticated tweeting devices that were able to provide geolocation capabilities via Global Positioning System (GPS) and/or cellular location methods [47-50]. The United States was oversampled to provide a suitable basis for exploratory correlations between city demographic and climatic data with pain-related tweet sentiment. Additional technical details are available in Multimedia Appendix 1.

The search was repeated for each of the 50 selected cities. A total of 10% of the tweets from each city were visually inspected for quality assurance. Data from one city were found to be corrupt, we believe, due to an error in our query code and were removed from further analysis. Given that all other tweets were collected in a batch search, we elected not to repeat the collection of this city's data given concerns for skewing of sampling due to different search times.

Of note, tweets in this analysis were not specifically searched for "#pain," whereby the hashtag is used as a metatag to mark a tweet as containing a specific topic [18]. We opted to search for "pain" as a general search term to discern how the term was used in the normal discourse of daily life. A search specifically for "#pain" would have returned only those tweets wherein the tweet's topic of interest was identified by the tweeter as pain, thus biasing the returned context and sentiment of the tweet contents.

Tweets obtained from this sample were consolidated into a pain tweet corpus, consisting of the text of all collected tweets. Here, "corpus" (and its plural, "corpora") refers to a body of texts on which analyses are conducted.

Term-Term Association Measurements With Graph Analysis

To measure how often terms in a tweet were associated with the term "pain" or other terms, we used an analytical approach known as graph analysis [6]. Each term was represented as a node in a network and the relationships between terms were the links, or edges, connecting those nodes. Note that in the content-analysis experiments, nodes represented individual words and communities represented groups of associated words connected by links or edges. Whenever 2 terms were found in the same tweet, those 2 terms were considered to share a link. The linkage of nodes by edges lends itself to quantitative analysis via those matrix algebra methods that underpin graph theory. Additional technical details are available in Multimedia Appendix 1.

For each term, the total degree centrality was first calculated by counting how many different links, or edges, that term had to other terms within the corpus. By examining how well groups of terms were associated with one another, but not other terms or groups of terms, communities of terms commonly associated with one another were determined using a community detection algorithm based on the Louvain method [3].

Sentiment Analysis

Sentiment scoring of tweets combined a rule-based approach with a statistical modeling system to create a hybrid sentiment classifier [51]. The rule-based approach used the AFINN (named for the author, Finn Årup Nielsen) listing of weighed positive and negative keywords [52]. The AFINN wordlist is a list of manually labeled English terms that have been rated for positive versus negative polarity, which has been explicitly validated for use in microblog environments such as Twitter. This was supplemented with emoticon terminology to enhance the accuracy of the rule-based classifier [53-55]. Additionally, the rule-based approach incorporated negation terms and contractions within 5 terms of a positive or negative keyword to reverse the sentiment to a score of ±1. By summing the positive and negative weights of keywords identified within a given tweet, the polarity (positive versus negative sentiment) could be calculated along with a confidence level. The statistical model employed a Naïve Bayes algorithm with a smoothed relative frequency for text normalization and a feature-ranking algorithm based on the risk ratio [48]. Additional technical details are available in Multimedia Appendix 1.

Classifier scores were compared with human ratings of sentiment using an interrater agreement scoring system. Given initial concerns over the implementation of sentiment analysis, each reviewer was engaged in a short didactic session by the principal investigator (PT) and given specific examples, including "Exercise was great! No pain, no gain!" for positive sentiment versus "Twisted ankle, pain unbearable, so sad to miss game!" for negative sentiment. However, given the subjective nature of sentiment analysis and exploratory nature of this characterization, more formal training was not offered. Given the historically poor interannotator agreement with sentiment analysis, some have suggested that the decidedly deterministic results provided by rule-based and classifier-based sentiment analyses may offer methodological advantages over those offered by human annotators [11,56,57].

Exploratory analyses correlated elementary demographic and climatic data for US cities with the proportion of pain-related tweets with positive sentiment for those cities. This exploratory analysis was motivated by historical clinical wisdom as well as work by Keller et al [58] and Jamison et al [59] that suggests an association between cooler climates, decreased mood, and greater pain intensity. Population, population density, median age, percentage of high school graduates, percentage with bachelor's degree or higher, median household income, and number of individuals below poverty level were obtained from the 2010 US Census [60]. Given its rural nature, data for Phoenix Township, Arkansas, were extracted from data pertaining to Pope County, Arkansas, in the absence of specific data from the US Census. Percentages of individuals without health insurance were extracted from the 2010 Small Area Health Insurance Estimates dataset on a per-county basis [61]. Climate data for the month of September for each city were aggregated from the 1981-2010 normals published by the National Oceanic and Atmospheric Administration [62], and included average high temperatures and average number of precipitation days with greater than 0.01 inches of rain. Climate data for Phoenix Township, Arkansas, was adapted from the Little Rock, Arkansas, climate area.

Community Structure of Twitter-Based Social Networks Related to Pain

Collection of Retweet Data

In March of 2013, we searched Twitter for the following terms: pain, #pain, happy, excitement, sad, fear, tired, anguish, apple, Manchester United, and Obama [63]. Emotional terms were selected as samples of positive and negative pain-related terms from a prior compilation of 15 universally applicable, cross-cultural emotional affects [39,40]. Comparator terms were empirically chosen to reflect discourse on common topics in an effort to compare against topics with widespread media attention across different public domains following discussion with coauthors. Each search was filtered for English-language tweets and was limited to 1500 returned tweets by the Twitter API. Additional technical details are available in Multimedia Appendix 1.

Description of Social Network Analysis Measurements

After import into Gephi, the network- and node-level metrics were calculated for each search term [8,64,65]. Network-level metrics included node and edge count, network diameter, average path length, density, and the number of weakly and strongly connected components [66,67]. Calculated node-level metrics included the number of modularity communities, the total degree centrality, in-degree centrality, and out-degree centrality [3,65].

To determine how often other emotion terms were tweeted by those individuals engaged within the pain retweet network, we sampled 100 individuals from the pain term network who submitted a tweet containing the term “pain” as a retweet or mention to another Twitter user. Using the userTimeline function (a specific piece of computer code within the twitteR package created for use with the R programming language) in the twitteR package, we then requested up to the last 100 tweets for each of these individuals. The text of these tweets was combined into a corpus. This corpus was then searched for the number of occurrences of each of the 6 emotion terms (happy, excitement, sad, fear, tired, anguish) and 3 objective terms (apple, Manchester United, Obama) previously noted. For each term, its frequency and its frequency in proportion to the frequency of the term “pain” were calculated and reported. Additional technical details are available in Multimedia Appendix 1.

Results

Content of Pain-Related Tweets

Graph Analysis

Analyses were conducted on a version of the pain tweet corpus in which identical tweets were removed; this is referred to as the reduced pain tweet corpus. For the graph analysis, the reduced pain tween corpus contained 47,958 nonduplicate tweets. The most common terms found within the reduced pain tween corpus included “feel” (n=1504), “don’t” (n=702), “love” (n=649), “can’t” (n=543), “ass” (n=374), “time” (n=340), “life” (n=328), “lol” (n=327), “hurt” (n=294), and “people” (n=288) (Multimedia Appendix 2). There were a total of 14,877 terms that were contained within the reduced pain tween corpus and these terms were connected across 451,209 edges.

The average degree centrality of the reduced pain tween corpus graph was 60.7, with total degree centrality counts for individual terms ranging from 0 to 5652 with a median of 18 (Figure 1). Terms with the highest total degree centrality included “feel” (degree centrality=5652), “don’t” (degree centrality=3375), “love” (degree centrality=3274), “ass” (degree centrality=3049), and “can’t” (degree centrality=2983) (Multimedia Appendix 3). The most common associations between terms, as a function of edge weights, included “laugh” and “watching” (edge weight=566), “don’t” and “feel” (edge weight=395), and “uploaded” and “video” (edge weight=361) (Table 1). A total of 161 modulus-based communities were detected using Louvain’s algorithm (Figure 2). The 10 most common modulus communities accounted for 77% of all terms.

Table 1. Edge weights of frequent associations between terms.

Rank	Term 1	Term 2	Edge weight
1	Laugh	Watching	566
2	Don’t	Feel	395
3	Uploaded	Video	361
4	(Name)	Laugh	335
5	Hart	Laugh	310
6	Feel	Lol	283
7	Feel	Love	276
8	Cant	Feel	222
9	Hart	Kevin	200
10	“ ”	Feel	183
11	Waking	Worst	171
12	Baby	Bring	166
13	Hope	Running	164
14	House	Running	163
15	Please	Running	161
16	Chicago	Running	160
17	Marathon	Running	160
18	Miles	Running	160
19	iPhone	Temper	158
20	Carriers	iPhone	158
21	Hope	iPhone	158
22	iPhone	Margin	158
23	(Name)	Watching	155
24	Carriers	Temper	147
25	Hope	Temper	147

Figure 1. Graph of reduced pain-related tweet corpus. Each term contained within corpus is represented by a point; point size corresponds to the total degree centrality of the associated term. The color of each point indicates membership to a modularity community. Whenever a term is associated with another term within a given tweet, the 2 points are connected by a line, or edge; edge width corresponds to the frequency of association between the 2 connected terms.

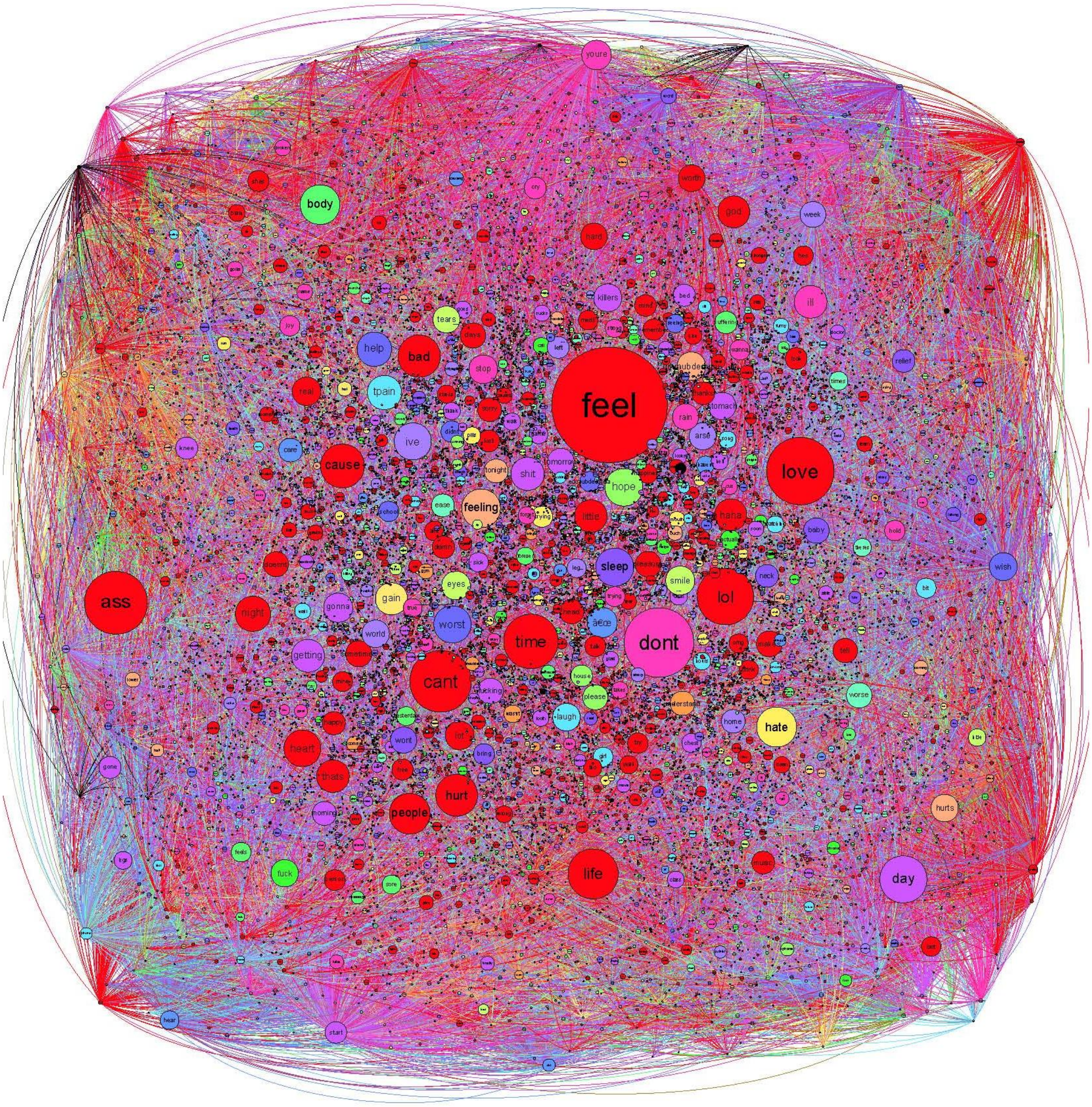
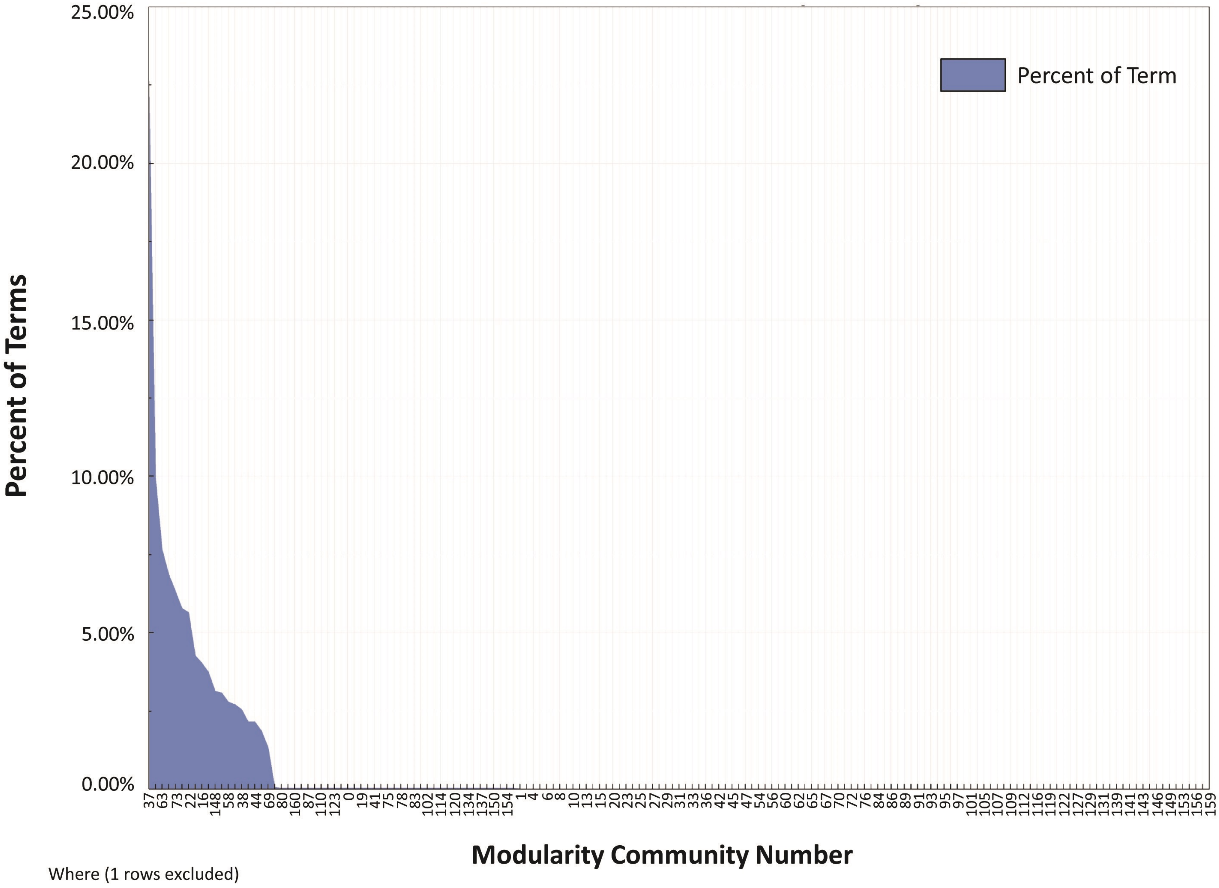


Figure 2. Percentage of terms contained within 161 modularity communities.



Sentiment Analysis

The sentiment classifier was validated in stages. In the first stage, the rule-based classifier, which was originally intended for classifying a broad array of text topics, was tested on 3 test sets: a 1500-tweet corpus based on a search for "happy," a 1500-tweet corpus based on a search for "sad," and a 1500-tweet corpus based on a search for "apple." The rule-based classifier identified 92.67% (1390/1500 of "happy" tweets as positive in sentiment, 19.53% (293/1500) of "sad" tweets as positive in sentiment, and 38.32% (575/1500) of "apple" tweets as positive in sentiment. The naïve Bayesian classifier, which was specifically trained on tweets pertaining to pain, identified 89.64% (1345/1500) of the "happy" tweets as positive in sentiment, 69.7% (1046/1500) of "sad" tweets as positive in sentiment, and 90.24% (1354/1500) of "apple" tweets as positive in sentiment (see Multimedia Appendix 4).

In the second validation stage, the sentiment classifier was tested on a holdout set of 100 tweets from the pain tween corpus not previously used for training of the naïve Bayesian component. When rated by humans, this test set contained 38% (38/100) (author PJT), 37% (37/100) (author RG), and 19% (19/100) (author MG) positive tweets depending on the rater, with a Cohen's kappa of .42, suggesting low to moderate interrater agreement. The rule-based component identified 42% (42/100) of these tweets as positive and the naïve Bayesian component identified 38% (38/100) as positive, with a Cohen's kappa between the 2 components of .16. When combined with the naïve Bayesian component to create the final hybrid classifier, a total of 39% (39/100) of the pain tween corpus test-set tweets were rated as positive in sentiment. The Cohen's kappa for the hybrid, rule-based, and naïve Bayesian classifier was .382, and for the human raters and the hybrid classifier was .317 (Multimedia Appendix 3).

Sentiment analysis was conducted on the entire pain tween corpus of 65,410 tweets. Sentiment scores of pain-related tweets were first compared among cities. The proportion of tweets with positive sentiment ranged from 13.13% (197/1500) in Manila, Philippines, to 55.73% (836/1500) in Los Angeles, California, with a median of 29% (Figure 3). There was a statistically significant difference in the proportion of pain-related tweets with positive sentiment among the 49 tested cities ($P<.001$).

Sentiment scores of pain-related tweets were compared across a 24-hour period (Multimedia Appendix 5). The proportion of tweets with positive sentiment ranged from 23.88% (833/3488) at 1600 to 38.25% (469/1226) at 2400, with a median of 32% (Figure 4). There was a statistically significant difference in the proportion of pain-related tweets with positive sentiment across the 24-hour period ($P<.001$).

Correlations between city-level demographic and climate characteristics and the percentage of pain-related tweets with positive sentiment were examined as an exploratory analysis (Table 2). Statistically significant correlations were observed between the percentage of positive tweets and the percentage of individuals without health insurance (ρ=.476, *P*=.02), average high temperature for September (ρ=.425, *P*=.03), and the latitude of the city (ρ=–.42, *P*=.04).

Table 2. Spearman rank correlations (ρ) between proportion of positive tweets and city-level demographic and climate data.

Variable	ρ	*P*
Percentage without health insurance	.476	.02
Average high temp in September	.425	.03
Latitude	–.420	.04
Longitude	–.358	.08
Average precipitation days in September	–.305	.14
% High school graduate	–.198	.35
% Bachelor’s degree or higher	.180	.40
Individuals below poverty level	–.169	.42
Median age	.166	.43
Population density	–.111	.60
Median household income	.108	.61
Population	–.018	.93

Figure 3. Percentage of pain-related tweets with positive sentiment in selected North American cities. Larger diameter circles indicate higher proportions of positive sentiment in tweets containing the term “pain.”.

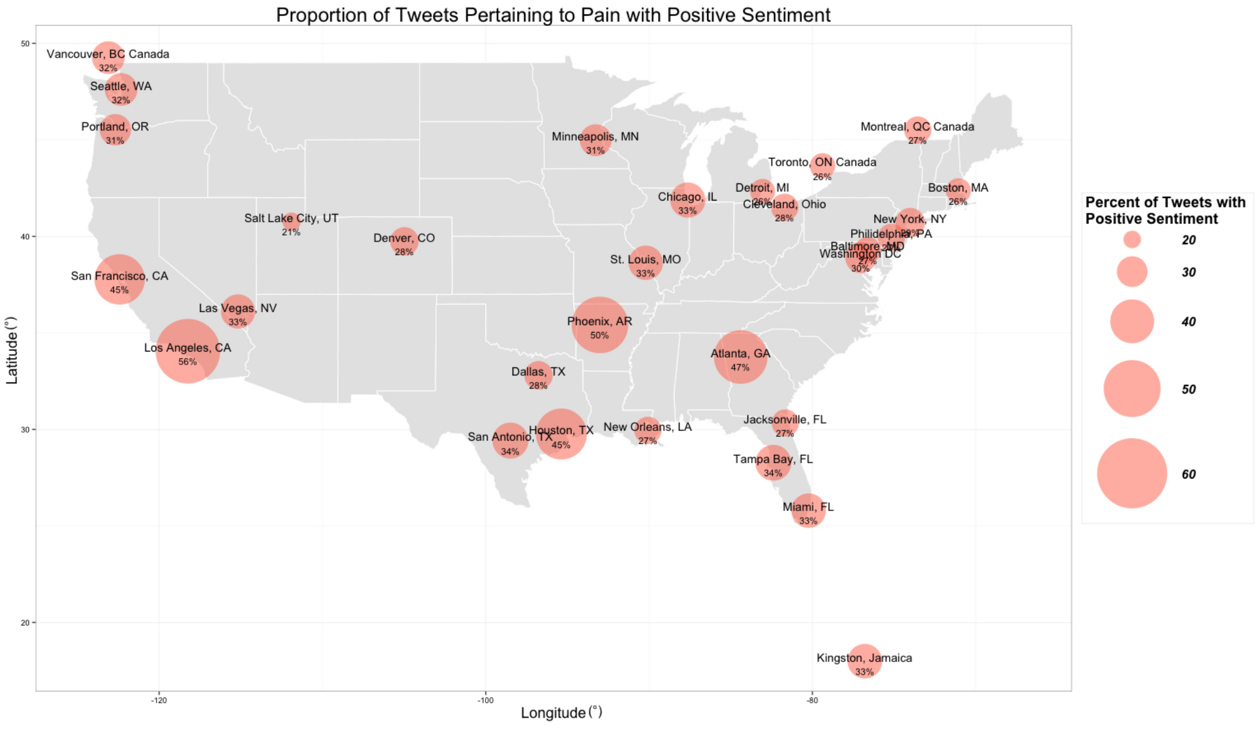
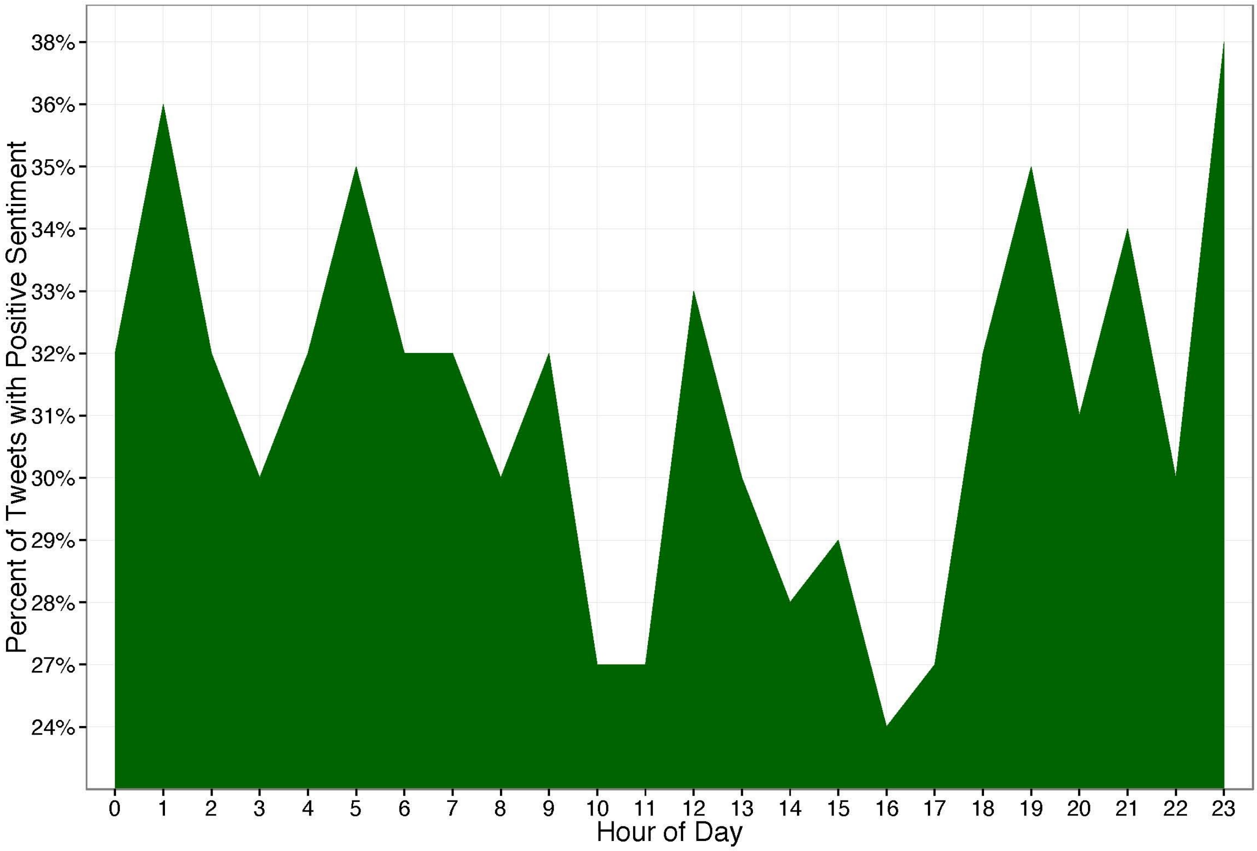


Figure 4. Percentage of pain-related tweets that contained date and time stamps with positive sentiment over a 24-hour period. Times were adjusted from UTC to local times according to geographic location.



Community Structure of Twitter-Based Social Networks Related to Pain

Of 16,500 tweets equally distributed across 11 search terms, 48.28 % (7967/16,500) were involved in retweet networks. By visual analysis of the retweet networks, those pertaining to “pain” exhibited greater sparsity and lower connectedness than did those pertaining to “apple,” “Manchester United,” and “Obama” (Figure 5). The Obama network had the greatest number of retweeting nodes (964), and the Manchester United network had the greatest number of edges (n=827) (Multimedia Appendix 6). Network diameter, average path length, and network density did not differ greatly between the compared networks. The number of weakly connected network components, in proportion to the total number of nodes, was greater for emotional terms when compared with specific objects, ranging from a minimum of 0.14 for Manchester United to between 0.37 for pain, 0.43 for happy, and 0.45 for tired. By contrast, the objective terms overall maintained the greatest percentage of their nodes within the giant component (Figure 6). Manchester United’s network maintained 47% percent of its nodes within the giant component, followed by 29% for Obama and 25% for apple. The emotional terms exhibited lower percentages at 9% for #pain, 4% for pain, 3% for sad, and 2% for happy. An important exception to this trend is the network for fear, which maintained 56% of its nodes within the giant component

Similar to the results for weakly connected network components, the number of modularity communities in proportion to node count was greater for emotional terms such as tired (0.45), happy (0.43), and sad (0.4) when compared with objective terms such as apple (0.26), Obama (0.25), and Manchester United (0.14) (Figure 7). Maximum in-degree centrality scores were greater than out-degree centrality for all terms, although the median numbers for all centrality scores remained between 0 and 1 for all terms (Multimedia Appendix 7). Maximum in-degree centrality scores were greater for objective terms in comparison with emotional terms. In particular, there were statistically significant differences between “apple” and “pain” (mean score difference=−65, *P*=.003, effect size=0.10), “excitement” and “pain” (mean score difference=−70, *P*=.001, effect size=0.10), “Manchester United” and “pain” (mean score difference=−167, *P*<.001, effect size=0.23), and “fear” and “pain” (mean score difference=−175, *P*<.001, effect size=0.23) for in-degree centrality. For out-degree centrality, there were statistically significant differences between “Manchester United” and “pain” (mean score difference=182, *P*<.0001, effect size=0.25), “fear” and “pain” (mean score difference=163, *P*<.001, effect size=0.21), “Obama” and “pain” (mean score difference=79, *P*=<.001 effect size=0.10), and “apple” and “pain” (mean score difference=65, *P*=.002, effect size=0.10). For total degree centrality, there were only statistically significant differences between “Obama” and “pain” (mean score difference=79, *P*<.001, effect size=0.13), and tired and pain (mean score difference=−37, *P*=.002, effect size=0.10) (Multimedia Appendix 8).

In examining the frequency of other emotional and objective terms from the 100 sampled retweeters in the pain term network, we first identified 5967 other tweets published by these individuals. Notably, the term “pain” was mentioned only 35 times within this sample corpus (Table 3). The term “happy” had more than a 2-fold increase in frequency compared with “pain,” and “sad” and “fear” were represented at rates of 86% and 69% of that of pain. Despite their more complicated retweet network structures, the terms for “apple” (3%), “Manchester United” (0%), and “Obama” (14%) were found at substantially lower frequencies in proportion to pain than were the emotional terms.

Table 3. Occurrences of emotive terms in 100-user sample of pain network tweeters.[a]

Term	Frequency	Frequency in proportion to pain
Pain	35	1
Happy	73	2.09
Excitement	1	0.03
Sad	30	0.86
Fear	24	0.69
Tired	10	0.29
Anguish	0	0.00
Apple	1	0.03
Manchester	0	0.00
Obama	5	0.14

[a] Sampled 100 users from pain term network who submitted a tweet containing “pain” and a retweet or mention to an individual. Requested up to 100 of the most recent tweets from each of these individuals. 5967 tweets collected. Searched all text for these terms.

Figure 5. Panel of retweet networks for (A) pain, (B) #pain, (C) happy, (D) excitement, (E) sad, (F) fear, (G) tired, (H) anguish, (I) apple, (J) Manchester United, and (K) Obama. Each circle indicates a node, or Twitter user, and each line connecting the circles represents an edge, or a mention of 1 user in the tweet of another. Each edge is directional in that it "points" from the originating Twitter user to the recipient Twitter user. Node size reflects the degree centrality of the node, line thickness reflects the number of connections between nodes, and color reflects the connectedness community of a node.

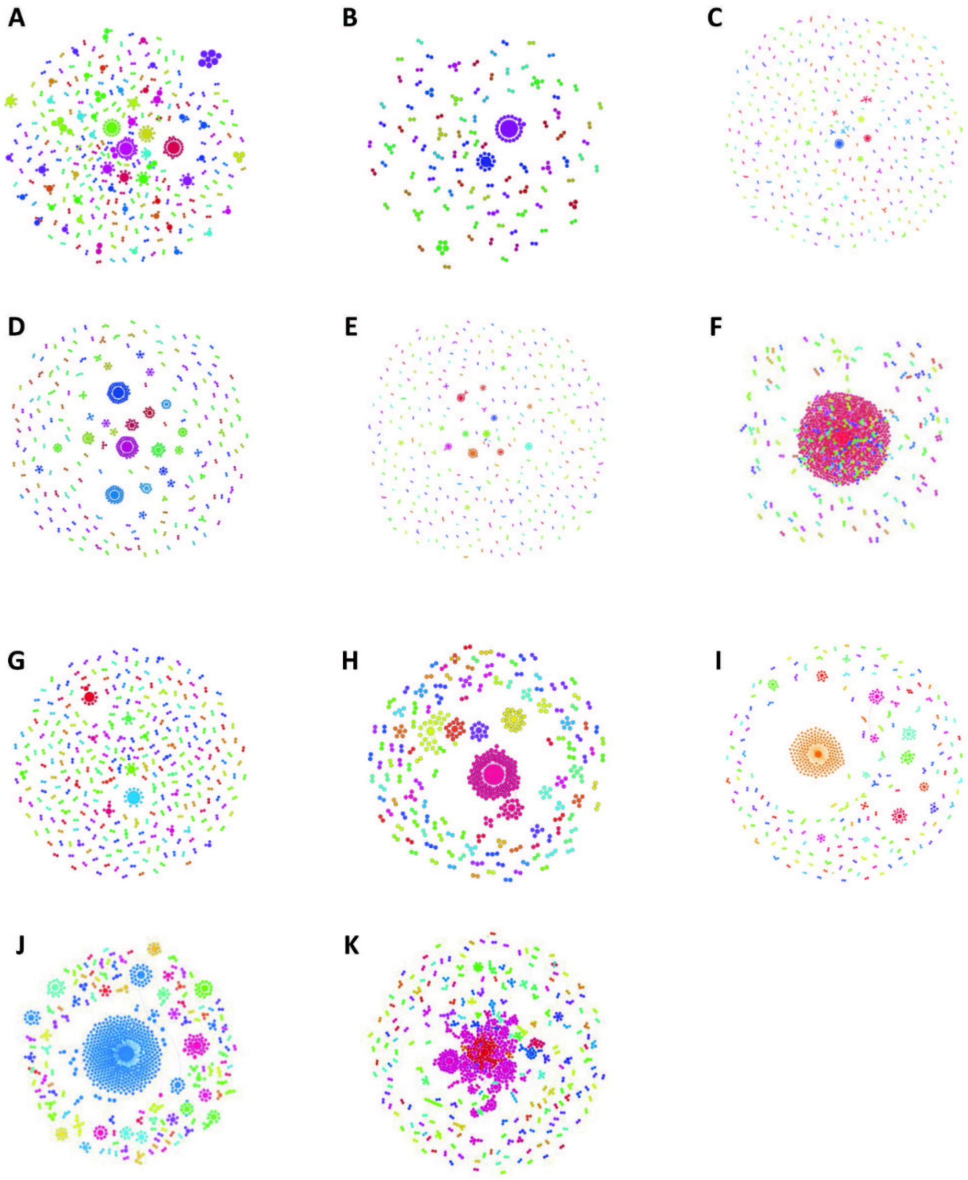


Figure 6. Total (blue) and giant component (red) nodes within retweet networks.

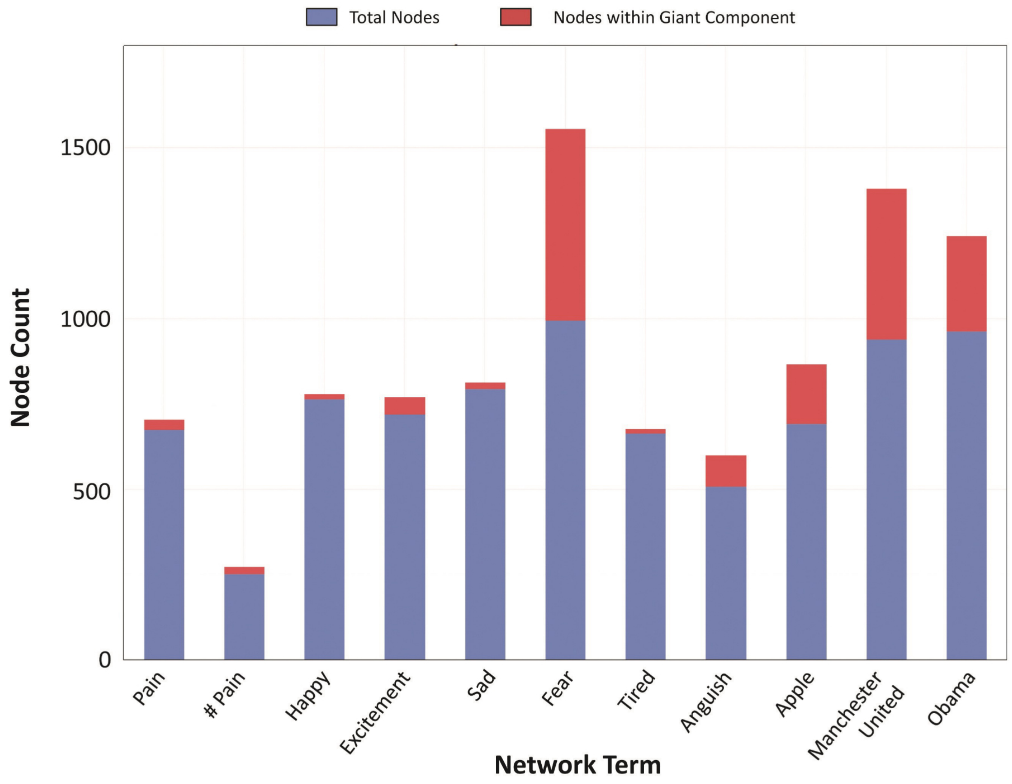
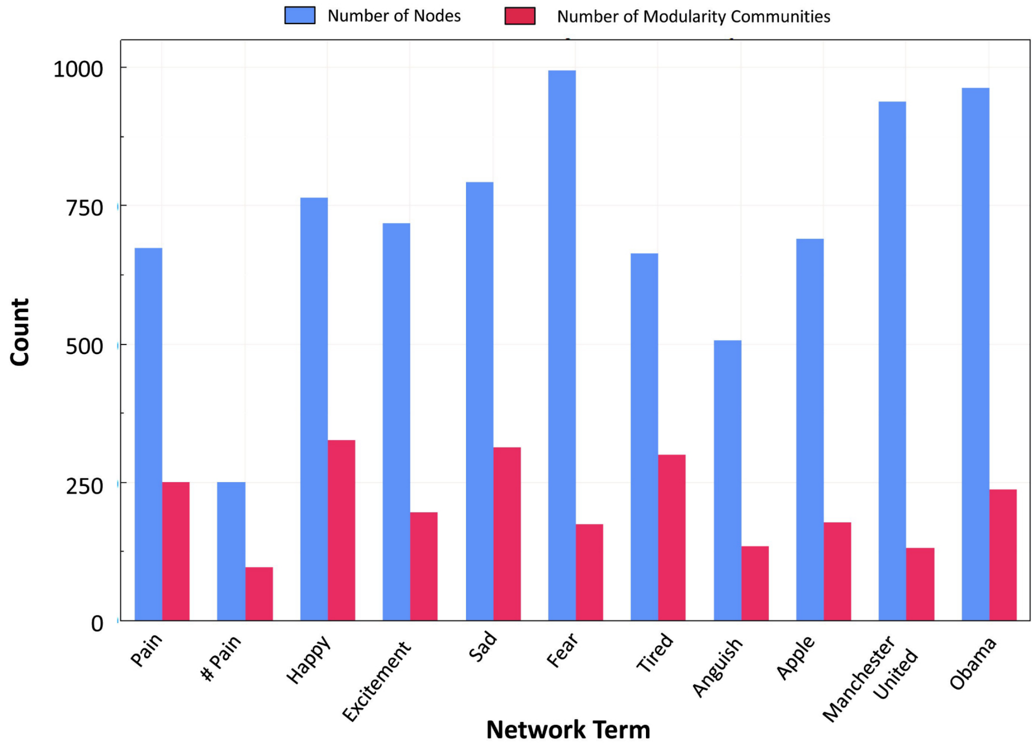


Figure 7. Number of nodes (blue) and modularity communities (red) per retweet network.



Discussion

The results presented here suggest that pain-related tweets carry special characteristics reflecting unique content and their communication among tweeters. The majority of tweets appear to describe pain within the context of relationships, although there were certainly a number of themes denoting specific physical pain. These data support the hypothesis that discussions of pain on Twitter do indeed focus on a range of both physical and nonphysical topics and not simply as a medical condition. Approximately one-third of pain-related tweets were quantified as containing a positive overall sentiment, a proportion that differed by geographic location and the hour of the day and supports our second hypothesis of a mixture of positive and negative sentiment in pain-related tweets. Our results also support the hypothesis of a unique connectivity pattern of retweet networks pertaining to pain.

Automated content analysis of pain-related tweets offers several potential applications to researchers, policymakers, and health care professionals. For instance, potential associations between biopsychosocial factors and tweet content may assist in the prediction of acute and chronic pain outcomes. More in-depth explorations of tweets related to pain may better differentiate physical versus emotional sources of pain through the use of extremely large datasets of tweets, although such differentiations would require calibration via external methods of data collection to ascribe the content to emotional versus physical sources with any degree of certainty. The volume of tweets available, coupled with their time and location tags, may permit analyses of seasonal and temporal shifts in pain density and their association with environmental and geopolitical events [22]. Indeed, prior work suggests that Twitter sentiment scores may lead public opinion surveys by a few days, thus offering an inexpensive way to indirectly assess public perception [68].

It may also be possible to use this method as an epidemiologic platform for ascertaining community health and a barometer of health care needs pertaining to pain in a manner similar to the experimental use of Twitter content as an influenza surveillance tool [69]. Such policy-based approaches to pain surveillance could help direct the allocation of pain management resources in time and location. Supplementing conduct of surveys of unmet pain needs on an annual or semiannual basis, the methods presented here could permit monthly, or even weekly, reviews of the effects of pain policy changes. Although limited by several critical factors, such as differentiation between acute and chronic pain, collection of data from a skewed population of social media users, and contextual relationships pertaining to pain, a streaming measure of the use and sentiment of emotive terms such as pain nevertheless may offer a low-cost, real-time supplement to such methods. Although such an initiative may only offer association-level data that may be of purely academic interest, it is at least feasible that calibration of such methods against established, robust (and sometimes cost-prohibitive) data collection measures may attach some value to such a social media-based approach to data collection, especially within targeted populations such as teenagers and Generation Y members. As suggested by Greaves et al [70], Twitter-based sentiment analyses could also help detect poor quality of pain-related health care delivery. This could lead to an efficient, widely deployable adjunct to the current Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) method of assessing patient satisfaction with pain control. Notably, existing work in this particular area points to widespread geographic disparities in patient experience with pain management during hospitalization, as well as agreement between social media ratings of hospitals and HCAHPS measures, thus pointing to an opportunity for social media sampling to supplement existing data collection methods [50,71].

Quantification of sentiment, especially as measured in a 140-character document rife with abbreviations and slang, poses valid questions concerning the accuracy and repeatability of the classifier. Prior work using related methods of sentiment quantification of tweets suggests that such sentiment analysis tracks well with important sociocultural events, although the magnitude of change in sentiment may be small and biased toward increases in negative rather than positive sentiment [16,29]. It is reassuring that such lexical accuracy has been demonstrated using word lists even without the inclusion of a machine learning component and has been reported for sentiment analysis projects targeting emotional constructs such as “happiness” [72].

Our analyses on correlations between the proportion of positive tweets and city-level demographics were intended to be exploratory given the lack of rigor in the selection of sampled cities. The correlations between average high temperature and latitude are reasonable given that higher temperatures and lower latitudes may be associated with greater sunlight exposure and more positive affect [58]. The presence of geographic differences in tweet sentiment is in keeping with very recent work comparing the “happiness” of cities in the United States by measuring the overall sentiment of 10 million geotagged tweets collected in 2011, although such comparisons are limited to the presence of differences given that this study did not examine the role of latitudes and temperatures in association with city ”happiness” [20]. When reviewing such matters, it is important to consider the risk of ecological fallacy given the simultaneous measures of individuals and groups.

Our results suggest that the sentiment of tweets pertaining to pain differs over the 24-hour interval. This is in keeping with prior work by Thelwall et al [29], which suggests that sentiment seems most negative in the late morning and late afternoons. However, this contrasts with work by Dodd et al [18], which showed small increases in positive sentiment between the hours of 0500 and 0600 and again between 1900 and 2000. Our results similarly show spikes in positive sentiment at 0500 and 2000, as well as supplemental spikes at 1200 and sustained increases after 2000. Interestingly, in the United States, the time intervals with lowest sentiment for pain-related tweets coincide with the hours that frequently precede mealtime.

One interesting observation noted in this exploratory analysis was the association between positive sentiment of pain-related tweets and areas with a high percentage of individuals without health insurance. It is possible that the skewed demographics of social media users may also be those without health insurance, which would be incongruent with prior data associating chronic pain with access to health services [73]. This may further point to the discrepancy of pain as a disease state versus the representation of the concept of pain within social media platforms.

Aside from content analysis, examination of pain-related tweets can also uncover information about the online social networks of those tweeters who tweet about pain as a matter of discourse. In recent years, multiple teams have explored social media platforms as they relate to social support systems for physical and mental health challenges [12,23,26,42-46]. Indeed, an automated social network analysis of pain-related tweets of patients may serve to quantify and monitor treatment progress for many potential patient-centered outcomes. Such approaches can be simultaneously applied to individual patients and entire communities, thus helping policymakers gauge the effectiveness of large-scale treatment interventions as well as provide decision support regarding resource allocation.

Retweet patterns for tweets pertaining to “pain” yield smaller discussion communities than do tweets on objective subjects. Those users involved in pain-related discussions were weakly connected through giant components of smaller size and were more likely to participate instead in one of a larger number of smaller modularity communities. Taken together, these results suggest that tweeters tend not to promote statements from others pertaining to pain, as they might with tweets on subjects such as sports or politics. Notably, this is not terribly dissimilar in structure to retweet networks for ”apple” and even ”Manchester United,” although the component sizes for pain retweets are indeed much smaller.

In contrast to prior work on the use of social media outlets for social support systems, our results suggest that such publications pertaining to pain may not trigger social media equivalents of “conversations” as they might if one were making a statement concerning sports or politics [12,23,26,41-46]. For pain and other emotional terms, most retweets were “dead end” expressions with a path length of 1. However, for objective term networks, retweets seemed to “echo” prior content across multiple successive communities, leading to longer path lengths. Tweeters expecting responses to tweets about pain may thus be surprised at a perceived lack of empathy from the Twitterverse.

Saito and Masuda [74] have demonstrated 2 types of popular tweeters: the first has many followers but follows only a small number of individuals themselves, whereas the second maintains large communities of followers and followed sources. The pain and #pain retweet networks mostly followed the type 1 schematic of Saito and Masuda, and the type 2 schematic predominated for the objective term networks. Others have differentiated retweet behaviors into “broadcasters,” or those with many followers but who follow few sources themselves, and “miscreants,” or those with few followers but who follow many sources. Here again, we see that pain-related tweets follow a miscreant pattern of retweets, whereas objective term networks, which echo earlier tweets into multiple communities, follow the broadcasters pattern of retweeting [9].

The lack of retweeting about pain may indeed limit the utility of Twitter, at least as a limited dataset, in studying pain-related discussions. On the other hand, the presence of tweets about pain in the context of an overall low prevalence on the topic may offer an important insight into a given tweeters’ focus on pain. To this end, earlier work on social media and chronic medical conditions suggests that alternative social network media, such as Facebook, contains more health care groups than can be found on Twitter [75]. Prior work with tweets pertaining to incontinence have noted a lack of “useful content,” suggesting that some medically oriented Twitter content lacks a suitable substrate for conversation [21]. Given the findings by Kumar et al [76] suggesting that topics of discussion heavily influence a user’s interest in participating in Twitter-based discussions, it may simply be that tweeters are uninterested in discussing topics related to pain. Regardless, the observed lack of social promotion of pain-related tweets may limit the utility of Twitter as a medium for promoting social interactions in those with impaired social networking due to chronic pain.

This work opens several interesting possibilities pertaining to pain research. The volume of tweets available, coupled with their time and location tags, may permit analyses of seasonal and temporal shifts in pain density and their association with environmental and geopolitical events [57]. Such policy-based approaches to pain surveillance could help direct allocation of pain management resources in time and location in a manner similar to that of Twitter-based resource allocation during natural disasters [4,16,17,23-28]. Instead of conducting surveys of unmet pain needs on an annual or semiannual basis, the methods presented here could permit monthly, or even weekly, reviews of the effects of pain policy changes. As suggested by Greaves et al [70], Twitter-based sentiment analyses could also help detect poor quality of pain-related health care delivery. However, such benefits must be viewed in the context of the complexity of the task of searching through large volumes of tweets to identify pain-related material and then processing this material into relevant information that can be used for research and/or decision support.

Given the scope of our project, we accumulated several limitations pertaining to methods available for Twitter-based research. First, given that Twitter is predominantly used by younger individuals more comfortable with technology, our results do not account for large swaths of the general population. A 2012 survey by the Pew Research Center suggests that 16% of Internet users use Twitter and that Twitter “is especially appealing to” adults aged 18 to 29 years, African-Americans, and urban residents [77]. However, this is also true for volunteer studies in pain research, which traditionally sample primarily from young adult populations [73]. Our study examined only public tweets; it is certainly possible that the preceding results may be skewed even among social media users given that social media users who tweet about their pain experiences may choose to keep their postings private due to the personal nature of this topic. Our work was similarly limited in its use of only English-language tweets; therefore, our data are likely not representative of all tweets originating from cities that are primarily non-English speaking. In examining groups of terms, our use of default community modularity coefficients may have led to inappropriately large populations of terms in the upper-tier communities. However, this approach also permitted encapsulation of broader topics, and minimized the chance of having topics populate multiple communities. The interrater kappa coefficients were admittedly low, but are in keeping with prior interrater annotation scores for sentiment analysis [11,56,57] that points to the subjective nature of sentiment analysis. The decision to study the selected cities was made empirically and was based on an effort to examine cities across a range of geographic regions and rural versus urban characteristics. Although a larger sample with a broader range of characteristics would have been attractive, this was unrealistic given that the size of the studied pain tween corpus grossly strained computing resources.

In conclusion, our results suggest that graph and sentiment analysis of pain-related tweets can offer important insights into the roles of pain throughout the social media discourse prevalent in today's society. Indeed, the preponderance of emotional and psychological pain references identified by our study suggests that future studies focusing on terms related to the physical manifestation of pain are necessary to explore this important aspect of pain research. Furthermore, the actual application of future semantic network analyses should include enhancements such as stemming, n-gramming, and synonym lists to improve the accuracy of their classifications. Further work is necessary to discern how geopolitical events and seasonal changes affect tweeters' perceptions of pain [37] and how such perceptions affect therapies for pain.

Acknowledgments

We would like to thank Corey Astrom for her considerable assistance in preparing this manuscript. This research was funded, in part, by a grant to Patrick J Tighe (NIH K23 GM 102697).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Pain tweet corpus generation.

[PDF File (Adobe PDF File), 88KB - jmir_v17i4e84_app1.pdf]

Multimedia Appendix 2

Most common terms in reduced pain tweet corpus.

[PDF File (Adobe PDF File), 38KB - jmir_v17i4e84_app2.pdf]

Multimedia Appendix 3

Terms with highest total degree centrality.

[PDF File (Adobe PDF File), 37KB - jmir_v17i4e84_app3.pdf]

Multimedia Appendix 4

Agreement statistics between human rates and classification methods for additional details of classifier performance, including specific information on sensitivity and specificity.

[PDF File (Adobe PDF File), 47KB - jmir_v17i4e84_app4.pdf]

Multimedia Appendix 5

Pain-related tweet volume by hour of day.

[PDF File (Adobe PDF File), 39KB - jmir_v17i4e84_app5.pdf]

Multimedia Appendix 6

Graph-level metrics.

[PDF File (Adobe PDF File), 41KB - jmir_v17i4e84_app6.pdf]

Multimedia Appendix 7

Node-level metrics.

[PDF File (Adobe PDF File), 45KB - jmir_v17i4e84_app7.pdf]

Multimedia Appendix 8

Effect sizes of in-degree, out-degree, and total degree centralities of retweet networks.

[PDF File (Adobe PDF File), 37KB - jmir_v17i4e84_app8.pdf]

References

1. https://business.twitter.com/en/basics/what-is-twitter/. 2014. What is Twitter? URL: https://business.twitter.com/twitter-101 [accessed 2014-05-22] [WebCite Cache ID 6PlSlpqgg]
2. Barber MJ. Modularity and community detection in bipartite networks. Phys Rev E Stat Nonlin Soft Matter Phys 2007 Dec;76(6 Pt 2):066102. [Medline: 18233893]
3. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct 09;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
4. 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]
5. Lancichinetti A, Fortunato S. Community detection algorithms: a comparative analysis. Phys Rev E Stat Nonlin Soft Matter Phys 2009 Nov;80(5 Pt 2):056117. [Medline: 20365053]
6. Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. Science 2009 Feb 13;323(5916):892-895. [doi: 10.1126/science.1165821] [Medline: 19213908]
7. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal Complex Systems 2006:1695.
8. Hawe P, Webster C, Shiell A. A glossary of terms for navigating the field of social network analysis. J Epidemiol Community Health 2004 Dec;58(12):971-975 [FREE Full text] [doi: 10.1136/jech.2003.014530] [Medline: 15547054]
9. Krishnamurthy B, Gill P, Arlitt M. A few chirps about Twitter. In: Proceedings of the first workshop on Online social networks. New York: ACM Press; 2008 Presented at: ACM SIGCOMM 2008; August 17-22, 2008; Seattle, WA p. 19-24. [doi: 10.1145/1397735.1397741]
10. Molloy M, Reed B. The size of the giant component of a random graph with a given degree sequence. Combinator Probab Comp 1998;7(3):295-305. [doi: 10.1017/S0963548398003526]
11. Nigam K, Hurst M. Towards a robust metric of opinion. In: Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications. 2004 Presented at: AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications; 2004; Stanford, CA.
12. Omurtag K, Jimenez PT, Ratts V, Odem R, Cooper AR. The ART of social networking: how SART member clinics are connecting with patients online. Fertil Steril 2012 Jan;97(1):88-94 [FREE Full text] [doi: 10.1016/j.fertnstert.2011.10.001] [Medline: 22088209]
13. Apfelbaum JL, Chen C, Mehta SS, Gan TJ. Postoperative pain experience: results from a national survey suggest postoperative pain continues to be undermanaged. Anesth Analg 2003 Aug;97(2):534-40, table of contents. [Medline: 12873949]
14. Connors AFJ, Dawson NV, Desbiens NA. The SUPPORT Principal Investigators. A controlled trial to improve care for seriously iii hospitalized patients: the study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). JAMA 1995;274:1591.
15. Committee on Advancing Pain Research, Care, and Education, Institute of Medicine. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education and Research. Washington, DC: The National Academies Press; Jun 29, 2011.
16. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. J Computat Sci 2011;2(1):1-8. [doi: 10.1016/j.jocs.2010.12.007]
17. Borondo J, Morales AJ, Losada JC, Benito RM. Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish Presidential election as a case study. Chaos 2012 Jun;22(2):023138. [doi: 10.1063/1.4729139] [Medline: 22757545]
18. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM. Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PLoS One 2011;6(12):e26752 [FREE Full text] [doi: 10.1371/journal.pone.0026752] [Medline: 22163266]
19. Melville P, Gryc W, Lawrence RD. Sentiment analysis of blogs by combining lexical knowledge with text classification. In: KDD '09 Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM Press; 2009 Presented at: 15th ACM SIGKDD international conference on Knowledge discovery and data mining; June 28-July 1, 2009; Paris, France. [doi: 10.1145/1557019.1557156]
20. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place. PLoS One 2013;8(5):e64417 [FREE Full text] [doi: 10.1371/journal.pone.0064417] [Medline: 23734200]
21. Prabowo R, Thelwall M. Sentiment analysis: a combined approach. J Informetr 2009;3:143-157. [doi: 10.1016/j.joi.2009.01.003]
22. Bollen J, Pepe A, Mao H. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Jul Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain.
23. Dufty N. Using social media to build community disaster resilience. Austral J Emerg Mgmt 2012;27(1):40-45.

24. Kumar S, Barbier G, Abbasi M, Liu H. Tweet tracker: an analysis tool for humanitarian and disaster relief. In: Fifth International AAAI Conference. 2011 Jul Presented at: AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain p. 1-2.
25. Manyez-Ortiz M, Albo-Portero C. Twitter in emergencies. Follow our roadmap. BMJ 2012;345:e4814. [Medline: 22826588]
26. Prochaska JJ, Pechmann C, Kim R, Leonhardt JM. Twitter=quitter? An analysis of Twitter quit smoking social networks. Tob Control 2012 Jul;21(4):447-449 [FREE Full text] [doi: 10.1136/tc.2010.042507] [Medline: 21730101]
27. 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]
28. Sullivan SJ, Schneiders AG, Cheang CW, Kitto E, Lee H, Redhead J, et al. 'What's happening?' A content analysis of concussion-related traffic on Twitter. Br J Sports Med 2012 Mar;46(4):258-263. [doi: 10.1136/bjsm.2010.080341] [Medline: 21406451]
29. Thelwall M, Buckley K, Paltoglou G. Sentiment in Twitter events. J Am Soc Inf Sci 2010 Dec 06;62(2):406-418. [doi: 10.1002/asi.21462]
30. Forgeron PA, McGrath P, Stevens B, Evans J, Dick B, Finley GA, et al. Social information processing in adolescents with chronic pain: my friends don't really understand me. Pain 2011 Dec;152(12):2773-2780. [doi: 10.1016/j.pain.2011.09.001] [Medline: 21963240]
31. Karp JF, Shega JW, Morone NE, Weiner DK. Advances in understanding the mechanisms and management of persistent pain in older adults. Br J Anaesth 2008 Jul;101(1):111-120 [FREE Full text] [doi: 10.1093/bja/aen090] [Medline: 18487247]
32. Lok EYC, Mok CC, Cheng CW, Cheung EFC. Prevalence and determinants of psychiatric disorders in patients with rheumatoid arthritis. Psychosomatics 2010;51(4):338-338.e8. [doi: 10.1176/appi.psy.51.4.338] [Medline: 20587762]
33. Maxwell SK, Barnett C, Kokokyi S, Leung JC, Yu JJ, Bril V, et al. Association of social support with quality of life in patients with polyneuropathy. J Peripher Nerv Syst 2013 Mar;18(1):37-43. [doi: 10.1111/jns5.12005] [Medline: 23521642]
34. Meldrum ML, Tsao JCI, Zeltzer LK. "I can't be what I want to be": children's narratives of chronic pain experiences and treatment outcomes. Pain Med 2009 Sep;10(6):1018-1034 [FREE Full text] [doi: 10.1111/j.1526-4637.2009.00650.x] [Medline: 19594848]
35. Saariaho TH, Saariaho AS, Karila IA, Joukamaa MI. Early maladaptive schemas in Finnish adult chronic pain patients and a control sample. Scand J Psychol 2011 Apr;52(2):146-153. [doi: 10.1111/j.1467-9450.2010.00849.x] [Medline: 21054422]
36. Soklaridis S, Cartmill C, Cassidy D. Biographical disruption of injured workers in chronic pain. Disabil Rehabil 2011;33(23-24):2372-2380. [doi: 10.3109/09638288.2011.573056] [Medline: 21504406]
37. Deter HC. Psychosocial interventions for patients with chronic disease. Biopsychosoc Med 2012;6(1):2 [FREE Full text] [doi: 10.1186/1751-0759-6-2] [Medline: 22293471]
38. Carr JL, Moffett JAK. The impact of social deprivation on chronic back pain outcomes. Chronic Illn 2005 Jun;1(2):121-129. [Medline: 17136918]
39. Moore CC, Romney AK, Hsia TL, Rusch CD. The universality of the semantic structure of emotion terms: methods for the study of inter- and intra-cultural variability. Amer Anthropol 1999;101:529-546. [doi: 10.1525/aa.1999.101.3.529]
40. Romney AK, Moore CC, Rusch CD. Cultural universals: measuring the semantic structure of emotion terms in English and Japanese. Proc Natl Acad Sci U S A 1997 May 13;94(10):5489-5494 [FREE Full text] [Medline: 9144265]
41. Heaivilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. J Dent Res 2011 Sep;90(9):1047-1051 [FREE Full text] [doi: 10.1177/0022034511415273] [Medline: 21768306]
42. Linnman C, Maleki N, Becerra L, Borsook D. Migraine tweets - what can online behavior tell us about disease? Cephalalgia 2013 Jan;33(1):68-69 [FREE Full text] [doi: 10.1177/0333102412465207] [Medline: 23136208]
43. McNeil K, Brna PM, Gordon KE. Epilepsy in the Twitter era: a need to re-tweet the way we think about seizures. Epilepsy Behav 2012 Feb;23(2):127-130. [doi: 10.1016/j.yebeh.2011.10.020] [Medline: 22134096]
44. Nicholson C. Palliative care on Twitter: who to follow to get started. Int J Palliat Nurs 2013 Jan;19(1):6-7. [doi: 10.12968/ijpn.2013.19.1.6] [Medline: 23354427]
45. O'Keeffe GS, Clarke-Pearson K, Council on Communications Media. The impact of social media on children, adolescents, and families. Pediatrics 2011 Apr;127(4):800-804 [FREE Full text] [doi: 10.1542/peds.2011-0054] [Medline: 21444588]
46. Sajadi KP, Goldman HB. Social networks lack useful content for incontinence. Urology 2011 Oct;78(4):764-767. [doi: 10.1016/j.urology.2011.04.074] [Medline: 21862115]
47. Cheng Z, Caverlee J, Lee K. You are where you tweet: a content-based approach to geo-locating Twitter users. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management. 2010 Oct 26-30 Presented at: 19th ACM International Conference on Information and Knowledge Management; October 26-30, 2010; Toronto, ON.
48. Hecht B, Hong L, Suh B, Chi EH. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. New York: ACM; 2011 May 7-12 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 07-12, 2011; Vancouver, BC.
49. Leetaru K, Wang S, Padmanabhan A, Shook E. Mapping the global Twitter heartbeat: the geography of Twitter. FM 2013 May 06;18(5):4366. [doi: 10.5210/fm.v18i5.4366]

50. Tighe PJ, Fillingim RB, Hurley RW. Geospatial analysis of Hospital Consumer Assessment of Healthcare Providers and Systems pain management experience scores in U.S. hospitals. Pain 2014;155:1016-1026.
51. Pang B, Lee L. Opinion mining and sentiment analysis. FNT in Information Retrieval 2008;2(1–2):1-135. [doi: 10.1561/1500000011]
52. Nielsen F. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In: Proceedings of the ESWC. 2011 May 30 Presented at: CEUR Workshop Proceedings; May 30, 2011; Heraklion, Crete.
53. Davidov D, Tsur O, Rappoport A. Enhanced sentiment learning using Twitter hashtags and smileys. In: Proceedings of the 23rd International Conference on Computational Linguistics. 2010 Aug 23-27 Presented at: 23rd International Conference on Computational Linguistics; August 23-27, 2010; Beijing, China p. 241-249.
54. Kouloumpis E, Wilson T, Moore J. Twitter sentiment analysis: the good, the bad, and the OMG!. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Jul 17-21 Presented at: AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain p. 1-4.
55. Liu KL, Li WJ, Guo M. Emoticon smoothed language models for Twitter sentiment analysis. In: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence. 2012 Jul 22-26 Presented at: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence; July 22–26, 2012; Toronto, ON.
56. Balahur A, Steinberger R. Rethinking sentiment analysis in the news: from theory to practice and back. In: Proceedings of WOMSA at the 2009 13th Annual Conference of the Spanish Association for Artificial Intelligence. 2013 Nov 13 Presented at: 13th Annual Conference of the Spanish Association for Artificial Intelligence; November 2009; Sevilla, Spain.
57. Bermingham A, Smeaton AF. A study of inter-annotator agreement for opinion retrieval. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2009 Jul Presented at: 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval; July 20-22, 2009; Dublin.
58. Keller MC, Fredrickson BL, Ybarra O, Côté S, Johnson K, Mikels J, et al. A warm heart and a clear head. The contingent effects of weather on mood and cognition. Psychol Sci 2005 Sep;16(9):724-731. [doi: 10.1111/j.1467-9280.2005.01602.x] [Medline: 16137259]
59. Jamison RN, Anderson KO, Slater MA. Weather changes and pain: perceived influence of local climate on pain complaint in chronic pain patients. Pain 1995 May;61(2):309-315. [Medline: 7659442]
60. United States Census 2010. URL: http://www.census.gov/2010census/ [accessed 2015-03-16] [WebCite Cache ID 6X50uni9G]
61. United States Census Bureau. 2014. Small Area Health Insurance Estimates URL: http://www.census.gov/did/www/sahie/data/interactive/index.html [accessed 2014-05-22] [WebCite Cache ID 6PlTKSYD8]
62. National Climactic Data Center, National Oceanic and Atmospheric Administration. 1981-2010 US Climate Normals URL: http://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals/1981-2010-normals-data [accessed 2015-03-16] [WebCite Cache ID 6X511XkLF]
63. Gentry J. twitteR: R Based Twitter Client. 2014. URL: http://cran.r-project.org/web/packages/twitteR/ [accessed 2014-05-22] [WebCite Cache ID 6PlStE6iN]
64. Borgatti SP, Everett MG. A graph-theoretic perspective on centrality. Social Networks 2006;28:466-484. [doi: 10.1016/j.socnet.2005.11.005]
65. Freeman LC. Centrality in social networks conceptual clarification. Social Networks 1978;1:215-239. [doi: 10.1016/0378-8733(78)90021-7]
66. Brandes U. A faster algorithm for betweenness centrality. J Math Sociol 2011;25:163-177. [doi: 10.1080/0022250X.2001.9990249]
67. Tarjan R. Depth-first search and linear graph algorithms. SIAM J Comput 1972;1:146-160.
68. O'Connor B, Balasubramanyan R, Routledge BR, Smith NA. From tweets to polls: linking text sentiment to public opinion time series. In: Proceedings of the International AAAI Conference on Weblogs and Social Media. 2010 Presented at: International AAAI Conference on Weblogs and Social Media; May 23-26, 2010; Washington, DC p. 122-129.
69. Paul MJ, Dredze M. You are what you tweet: analyzing Twitter for public health. Artificial Intelligence 2011:265-272.
70. 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]
71. 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 [FREE Full text] [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]
72. Dodds PS, Danforth CM. Measuring the happiness of large-scale written expression: songs, blogs, and presidents. J Happiness Stud 2010;11:441-456. [doi: 10.1007/s10902-009-9150-9]
73. Helme RD, Gibson SJ. Pain in older people. In: Crombie IK, editor. Epidemiology of Pain. Seattle, WA: IASP Press; 1999.
74. Saito K, Masuda N. Two types of Twitter users with equally many followers. PLOS ONE 2014;9(1):e84265.
75. de la Torre-Díez I, Díaz-Pernas FJ, Antón-Rodríguez M. A content analysis of chronic diseases social groups on Facebook and Twitter. Telemed J E Health 2012;18(6):404-408. [doi: 10.1089/tmj.2011.0227] [Medline: 22650380]

76. Kumar S, Morstatter F, Zafarani R, Liu H. Whom should I follow? Identifying relevant users during crises. In: HT '13 Proceedings of the 24th ACM Conference on Hypertext and Social Media. 2013 Presented at: 24th ACM Conference on Hypertext and Social Media; May 1-3, 2013; Paris, France p. 139-147.
77. Duggan M, Brenner J. The Demographics of Social Media Users - 2012. Washington, DC: Pew Internet & American Research Project; 2013 Feb 14. URL: http://www.pewinternet.org/2013/02/14/the-demographics-of-social-media-users-2012/ [accessed 2014-05-22] [WebCite Cache ID 6PlThW8dC]

Abbreviations

API: application programming interface
GPS: Global Positioning System
HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems

Edited by G Eysenbach; submitted 22.08.14; peer-reviewed by G Moseley, C Lalloo, L Jibb; comments to author 05.11.14; revised version received 17.12.14; accepted 21.01.15; published 02.04.15.

Please cite as:
Tighe PJ, Goldsmith RC, Gravenstein M, Bernard HR, Fillingim RB
The Painful Tweet: Text, Sentiment, and Community Structure Analyses of Tweets Pertaining to Pain
J Med Internet Res 2015;17(4):e84
URL: http://www.jmir.org/2015/4/e84/
doi:10.2196/jmir.3769
PMID:25843553

Original Paper

A New Source of Data for Public Health Surveillance: Facebook Likes

Steven Gittelman[1], PhD; Victor Lange[1], MS; Carol A Gotway Crawford[2], PhD; Catherine A Okoro[3], MSc, PhD; Eugene Lieb[4], MS, MBA, PhD; Satvinder S Dhingra[5], MPH; Elaine Trimarchi[1], BSc

[1] Mktg, Inc, East Islip, NY, United States

[2] USDA National Agricultural Statistics Service, Research and Development Division, Washington, DC, United States

[3] National Center for Chronic Disease and Health Promotion, Division of Population Health, Center for Disease Control and Prevention, Atlanta, GA, United States

[4] Custom Decision Support, Los Angeles, CA, United States

[5] Northrop Grumman, Atlanta, GA, United States

Corresponding Author:
Steven Gittelman, PhD
Mktg, Inc
200 Carleton Avenue
East Islip, NY, 11730
United States
Phone: 1 6314666604
Fax: 1 6312777601
Email: Steve@Mktginc.com

Abstract

Background: Investigation into personal health has become focused on conditions at an increasingly local level, while response rates have declined and complicated the process of collecting data at an individual level. Simultaneously, social media data have exploded in availability and have been shown to correlate with the prevalence of certain health conditions.

Objective: Facebook likes may be a source of digital data that can complement traditional public health surveillance systems and provide data at a local level. We explored the use of Facebook likes as potential predictors of health outcomes and their behavioral determinants.

Methods: We performed principal components and regression analyses to examine the predictive qualities of Facebook likes with regard to mortality, diseases, and lifestyle behaviors in 214 counties across the United States and 61 of 67 counties in Florida. These results were compared with those obtainable from a demographic model. Health data were obtained from both the 2010 and 2011 Behavioral Risk Factor Surveillance System (BRFSS) and mortality data were obtained from the National Vital Statistics System.

Results: Facebook likes added significant value in predicting most examined health outcomes and behaviors even when controlling for age, race, and socioeconomic status, with model fit improvements (adjusted R^2) of an average of 58% across models for 13 different health-related metrics over basic sociodemographic models. Small area data were not available in sufficient abundance to test the accuracy of the model in estimating health conditions in less populated markets, but initial analysis using data from Florida showed a strong model fit for obesity data (adjusted R^2=.77).

Conclusions: Facebook likes provide estimates for examined health outcomes and health behaviors that are comparable to those obtained from the BRFSS. Online sources may provide more reliable, timely, and cost-effective county-level data than that obtainable from traditional public health surveillance systems as well as serve as an adjunct to those systems.

(J Med Internet Res 2015;17(4):e98) doi:10.2196/jmir.3970

KEYWORDS

big data; social networks; surveillance; chronic illness

Introduction

The development of the Internet and the explosion of social media have provided many new opportunities for health surveillance. The use of the Internet for personal health and participatory health research has exploded, largely due to the availability of online resources and health care information technology applications [1-8]. These online developments, plus a demand for more timely, widely available, and cost-effective data, have led to new ways epidemiological data are collected, such as digital disease surveillance and Internet surveys [8-25]. Over the past 2 decades, Internet technology has been used to identify disease outbreaks, track the spread of infectious disease, monitor self-care practices among those with chronic conditions, and to assess, respond, and evaluate natural and artificial disasters at a population level [6,8,11,12,14,15,17,22,26-28]. Use of these modern communication tools for public health surveillance has proven to be less costly and more timely than traditional population surveillance modes (eg, mail surveys, telephone surveys, and face-to-face household surveys).

The Internet has spawned several sources of big data, such as Facebook [29], Twitter [30], Instagram [31], Tumblr [32], Google [33], and Amazon [34]. These online communication channels and market places provide a wealth of passively collected data that may be mined for purposes of public health, such as sociodemographic characteristics, lifestyle behaviors, and social and cultural constructs. Moreover, researchers have demonstrated that these digital data sources can be used to predict otherwise unavailable information, such as sociodemographic characteristics among anonymous Internet users [35-38]. For example, Goel et al [36] found no difference by demographic characteristics in the usage of social media and email. However, the frequency with which individuals accessed the Web for news, health care, and research was a predictor of gender, race/ethnicity, and educational attainment, potentially providing useful targeting information based on ethnicity and income [36]. Integrating these big data sources into the practice of public health surveillance is vital to move the field of epidemiology into the 21st century as called for in the 2012 US “Big Data Research and Development Initiative” [19,39].

Understanding how big data can be used to predict lifestyle behavior and health-related data is a step toward the use of these electronic data sources for epidemiologic needs [36,40]. Facebook has been used by individuals and public health researchers for novel surveillance applications [13,37,38,41-44]. For example, Chunara et al [13] used Facebook to examine the association between activity- and sedentary-related likes and population obesity prevalence. These researchers found that populations with higher proportions of activity-related Facebook likes had a lower prevalence of being overweight and/or obese. Facebook likes are a means by which Facebook users can identify their own preferred Internet sites and interests. Although Facebook likes are not explicitly health-related, researchers have shown that when taken together, the “network” of an individual’s likes are predictive of sociodemographic characteristics, health behaviors, obesity, and health outcomes [13,37,42,44]. Timian et al [44] examined whether Facebook likes for a hospital could be used to evaluate 2 quality measures (ie, 30-day mortality rates and patient recommendations) both quickly and inexpensively. Facebook likes have also been shown to be predictors of a variety of user attributes, such as intelligence, happiness, race, religious and political views, sexual orientation, and a spectrum of personality traits [37]. Researchers have proposed that Facebook likes be used as a new behavioral measure in a fashion similar to traditional questionnaires [37].

In this study, we focused on harnessing the predictive power of Facebook likes for enhancing population health surveillance. Toward this end, we viewed Facebook likes as a class of big data that may help us understand population health at a local level. Given that risk factors and associated health outcomes are often clustered in populations geographically [10,45,46], the ability to identify, monitor, and intervene at a community level exists. Although past research has used specific categories of likes to target theoretically related conditions (eg, [13]), it is possible that the entirety of the Facebook dataset can be used to form a complete profile of individuals that can be broadly applied to predictive models in a number of areas. If the Facebook characteristics of a region can predict physical activity, smoking, and self-management of chronic disease, then a strong argument can be made in favor of using these data to target, monitor, and intervene on adverse lifestyle behaviors.

In this paper, we examine how big data might be used to complement traditional surveillance systems. We explored the use of Facebook likes as potential predictors of health outcomes and the behavioral determinants of poor health outcomes at the county level. Specifically, we hypothesized that (1) Facebook likes provide a means of predicting county-level mortality, (2) Facebook likes can be used as an indicator of chronic disease outcomes (obesity, diabetes, and heart disease) that contribute to increased mortality, and (3) Facebook likes can be used as an indicator of adverse lifestyle behaviors that impact disease. If these hypotheses hold, then Facebook likes could ultimately be used to enhance population health surveillance.

Methods

Data Sources

Data for the analysis were collected from 4 sources. Objective reports on key health indicators (ie, life expectancy, mortality, and low birth weight) were collected from the National Vital Statistics System (NVSS) for 2011, which provides population data on deaths and births in the United States. According to its website, “these data are provided through contracts between [National Center for Health Statistics] NCHS and vital registration systems operated in the various jurisdictions legally responsible for the registration of vital events—births, deaths, marriages, divorces, and fetal deaths” [47].

Self-reported health outcome and risk behavior data were obtained from the Behavioral Risk Factor Surveillance System (BRFSS) [48]. The BRFSS is an ongoing random digit-dialed telephone survey operated by state health agencies with assistance from the Centers for Disease Control and Prevention (CDC). The surveillance system collects data on many of the behaviors and conditions that place adults aged ≥18 years at risk for chronic disease, disability, and death. The large sample size of the 2011 BRFSS (N=506,467) facilitated the calculation of reliable estimates for 214 counties with 500 or more respondents. In addition, the 2010 BRFSS facilitated the calculation of reliable estimates for 91% of counties in Florida—a year in which 61 of its 67 counties had 500 or more respondents. County-level risk factor data were obtained from the 2011 Selected Metropolitan/Micropolitan Area Risk Trends (SMART) BRFSS [49].

Facebook likes data were collected using the Facebook advertising application program interface (API) [50] in February 2013, which aggregates the number of users by zip code who expressed a positive inclination (“like”) toward certain categories of items by zip code. These zip code data were aggregated to the county level to allow for direct comparisons to the health data, with zip codes crossing borders assigned to the county they predominantly rest in. The data reflect the cumulative total of Facebook users’ likes at the time they were drawn. Out of 8 supercategories of available Facebook likes (ie, events, family status, job status, activities, mobile device owners, interests, Hispanic, and retail and shopping), 3 were deemed as potentially correlated with health and were selected for the model. The selected likes were activities, interests, and retail and shopping. These supercategories were selected because they contained items with an explicit theoretical relationship to health. For example, “interests” contains the “health and well-being” category, to which the relationship of health is self-explanatory. The “activities” category was chosen because it included “outdoor fitness and activities,” which seemed directly applicable to measures of physical activity, whereas “retail and shopping” was chosen due to its apparent linkage to socioeconomic status, a powerful driver of health outcomes (Multimedia Appendix 1) [51,52].

All constituent elements of these supercategories were used, regardless of a clear relationship to health, because the exact contents and means of construction of these data are not reported by Facebook. Other supercategories lacked these explicit links, although we acknowledge the possibility that potentially powerful indirect relationships may exist. Due to rounding performed automatically by the API that routinely led to overestimates, counties with fewer than 1000 profiles overall were excluded from the analysis. Facebook likes for each category were scored as a percentage of completed profiles in an area. Finally, to reduce multicollinearity caused by variation in levels of Facebook usage by county, values were divided by the average percentage of likes across all categories. The resulting variables can be characterized as a measure of popularity for each category relative to that of other categories. Although the individual variables resulting from this transformation were sometimes entirely uncorrelated with the originals, estimates using the raw and transformed variables correlated at R=.9. Thus, we concluded that the results of the proceeding analyses were not an artifact of this transformation.

Population data, such as average income, median age, and sex ratio, were collected using the 2010 US Census [53] and broken into county aggregates. Supporting county-level statistics unrelated to health were collected using “USA Counties Information” provided by the Census Bureau [54]. Overall, 214 counties in the continental United States contained sufficient data on all variables in the analysis.

Variables of Interest

Several sociodemographic, health outcome, and risk factor variables were selected for analysis. These included income, age, education, employment, nonwhite population, obesity, diabetes, physical activity, and smoking, as well as other measures such as general health status. A comprehensive listing, as well as the data source and assessment of each variable of interest are available in Multimedia Appendix 2.

Data Analysis

We began by using principal components analysis on the 37 Facebook likes categories within the 3 selected supercategories as a data reduction technique. We then used these factors in an ordinary least squares (OLS) regression to determine whether Facebook likes could predict a number of health outcomes, conditions, and related behaviors. Finally, by limiting our analysis to Florida, where available data were more comprehensive, we formed a predictive model via bootstrap regression [55] that demonstrated the predictive accuracy of Facebook in a visual format.

Results

The first stage in the analysis was to establish that health outcomes could indeed be determined by Facebook likes. Through principal components analysis, the 37 categories were reduced to 9 factors (varimax rotation) purely as a means of simplifying modeling efforts by reducing these categories into the latent sociobehavioral dimensions we believed they represented. This number was arrived on by applying the Cattell scree test (shown in Multimedia Appendix 3) [56], which evaluates the “elbow” in the distribution of eigenvalues; that is, the point at which additional factors do not seem to provide a substantial gain in variance explained. Each factor is numbered in accordance with the amount of variance it explains (Multimedia Appendix 4). Any attempt to interpret the actual nature of these factors is subject to errors in the interpretation of the Facebook advertising data; as such, we avoided the urge to do so. However, the factor loadings of each of the categories can be seen in Multimedia Appendix 5.

To test our hypothesis that Facebook likes can be used to predict mortality on their own, we used OLS regression. We used the 9 Facebook factors to predict life expectancy, with no other controls included in the initial model. The results, as shown in the “Facebook only” column of Table 1, were quite strong (model adjusted $R^{2=}$.69). Despite this relationship, Facebook only has value insofar as it provides predictive value beyond that of reliable data that is already available through the census or other means. Regression results for an OLS model predicting life expectancy with demographic information (average age and nonwhite population) and socioeconomic status (SES; as represented by average household income, unemployment rate, and percentage with bachelor’s degree) are shown in the “SES only” column of Table 1. There is a very strong relationship to be found there as well, although it is less strong than for Facebook factors alone. Finally, the 2 groups of variables are combined in the last column of Table 1, indicating that although a great deal of the variance in life expectancy is shared by both the Facebook and SES variables, the addition of Facebook improves the model fit above and beyond readily available socioeconomic measures. The resulting adjusted R^2=.81 also indicates that a considerable amount of the variation in county-level life expectancy can be explained by SES and Facebook likes.

Table 1. Ordinary least squares regression coefficients (β) for life expectancy (all independent variables are standardized).

	Facebook only		SES only		Facebook and SES	
	β	*P*	β	*P*	β	*P*
Facebook factor						
1	–0.14	<.001	—	—	0.20	<.001
2	0.79	<.001	—	—	0.43	<.001
3	–0.96	<.001	—	—	–0.30	<.001
4	0.60	<.001	—	—	0.42	<.001
5	0.69	<.001	—	—	0.41	<.001
6	0.21	<.001	—	—	–0.04	.05
7	–0.08	<.001	—	—	–0.04	.04
8	–0.61	<.001	—	—	–0.49	<.001
9	0.12	<.001	—	—	0.10	.70
Age	—	—	0.16	<.001	0.01	.87
Income	—	—	0.62	<.001	0.59	<.001
Education	—	—	0.88	<.001	0.61	<.001
Unemployment	—	—	–0.05	0.07	0.01	.70
Nonwhite population	—	—	–0.85	<.001	–0.47	<.001
Constant	77.08	<.001	77.06	<.001	77.06	<.001
Adjusted R^2	.69		.64		.81	
RMSE	1.28		1.29		1.01	

Table 2 summarizes regressions using the same set of predictors run across an array of health-related dependent variables and indicates the percent improvement in variance explained by the inclusion of Facebook likes when added to SES compared to the SES alone. There are 2 conclusions we can draw from this model. First, Facebook likes and SES in tandem prove to be effective predictors of all tested disease outcomes. Second, there is a persistent benefit of Facebook likes beyond that contributed by SES, although its magnitude varies widely.

Our third hypothesis posited that Facebook likes, as a measure of behavior, should be able to determine the behaviors that drive health outcomes. The results in Table 2 clearly show that Facebook likes had a sizeable impact in the predictive models of all tested health-related behaviors and in some cases, such as health insurance and exercise, the total model fit was quite strong.

Table 2. Facebook likes impact on model fit for 214 counties.

Dependent variable	Source[a]	Facebook, R^2	SES, R^2	SES + Facebook, R^2	Improvement with Facebook, %
Life expectancy	NVSS	.69	.64	.81	27%
Mortality	NVSS	.57	.49	.60	22%
Low birthweight	NVSS	.53	.17	.57	235%
Obesity	BRFSS	.46	.56	.60	7%
Diabetes	BRFSS	.36	.39	.55	41%
Heart attack	BRFSS	.32	.46	.46	0%
Stroke	BRFSS	.27	.30	.41	46%
Exercise	BRFSS	.57	.51	.76	49%
Insured	BRFSS	.48	.37	.65	76%
Self-Reported health	BRFSS	.51	.20	.55	175%
Smoker	BRFSS	.40	.42	.54	29%
Last checkup	BRFSS	.69	.30	.72	140%
Declined treatment	BRFSS	.39	.35	.49	40%

[a] BRFSS: Behavioral Risk Factor Surveillance System; NVSS: National Vital Statistics System.

Predicting Health Conditions

The natural extension of these findings would be to map out predicted prevalence of health conditions in data-deficient counties. Although 214 counties were sampled sufficiently for the BRFSS to provide county-specific estimates, the remaining 2895 counties were not. An additional source of data, such as Facebook, would be a cost-effective way to augment existing state-level data sources that are used to produce county-level estimates, such as the BRFSS.

However, attempting to apply predictions nationally from the 2011 SMART data creates a problem. Although predictions correlate well with actual levels in non-SMART data, mean levels are consistently upwardly biased. We hypothesized that the selection method that leads counties to be weighted according to the SMART program creates a nonrepresentative sample with better levels of general health than we see in the United States in general, particularly in areas that are more rural. As an alternative without such problematic selection issues, we limited our predictive model to 2010 Florida data. Florida collects more than 500 interviews in 61 of its 67 counties every 3 years, leading to a dataset that has neither sample size shortages nor selection biases relative to the state at large.

Using data exclusively from one state creates its own problems for a predictive model. Although the integrity of the data is very good, there is no easy way to correct for the various cultural differences between Florida and other states. Attempting to apply Florida-based models to the full set of SMART counties results in only fair level of correlation (R=.63). Although it indicates that relationships exist, this is not a sufficient level of accuracy on which to base policy decisions. Instead, we have limited our analysis to Florida to demonstrate the level of accuracy we feel can be achieved at a national level once a somewhat more representative selection of county-level data are made available.

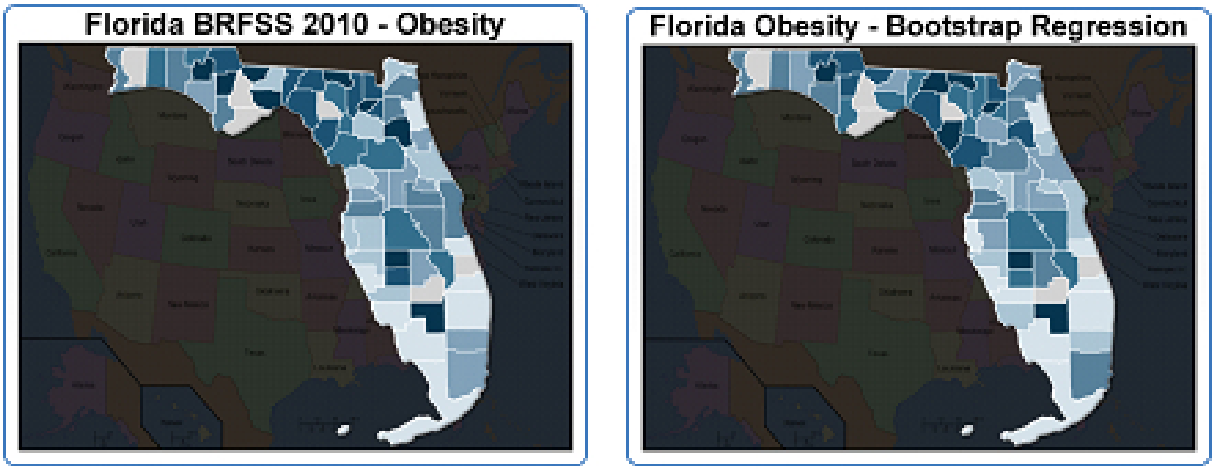
The results of a predictive model are shown in Table 3. These are the results of a bootstrap regression procedure in which 50 observations were drawn over 100 replications. Standard errors are high due to the limited sample size, but 2 of our Facebook likes categories retain their significance in the model. Although we would expect demographics and socioeconomic data to be very effective at predicting “healthy” versus “unhealthy” communities, we believe that the additional information provided by Facebook likes should help to clarify the finer distinctions between communities with similar general levels of health.

Table 3. Ordinary least squares regression (β) results for prediction of obesity.

Header	Facebook only		SES only		Facebook and SES	
	β	*P*	β	*P*	β	*P*
Facebook factor						
1	0.04	.05	—	—	–0.03	<.001
2	–0.02	.06	—	—	–0.01	.14
3	0.03	<.001	—	—	–0.01	.07
4	–0.02	.06	—	—	–0.01	.74
5	–0.02	.04	—	—	0.03	.01
6	–0.02	.07	—	—	–0.02	.13
7	–0.05	.30	—	—	0.02	.04
8	0.01	.34	—	—	0.01	.90
9	0.02	.36	—	—	–0.01	.17
Age	—	—	–0.01	.01	–0.01	.01
Income	—	—	–0.01	.37	–0.01	.59
Education	—	—	–0.03	<.001	0.01	.35
Unemployment	—	—	–0.01	.04	0.01	.58
Nonwhite population			0.02	.04		
Constant	0.29	<.001	0.30	<.001	0.30	<.001
Adjusted R^2	.77		.72		.8	
RMSE	0.03		0.03		0.03	

Figure 1 shows a graphical comparison of predicted values from the bootstrap regression procedure versus source data for obesity in Florida, where nearly all counties were sufficiently sampled for reliable estimates. These maps are dynamically shaded from light to dark in accordance with the level of obesity, with data separated into septiles of prevalence. As should be apparent visually, the fit is generally good—90% of errors in the model fall inside of ±2.1% (0.4 standard deviations) from Florida's estimated values from the 2010 BRFSS.

Figure 1. Actual statistics compared with predicted values for obesity, 2010 BRFSS. Darker colors represent higher prevalence. Light gray indicates missing data.



Discussion

When we first undertook this research plan, it was our expectation that the larger part of the measurement error that would affect our results would come through the imprecise categorization and geographic aggregation of the Facebook data. However, although there are some exceptions, the consistency and strength of fit we have found seem manifest. Our models do extremely well in predicting levels of health variables across counties where data are plentiful, and often diverge from BRFSS estimates where they are not. This suggests the possibility that data imputed from Facebook and vital statistics may provide a more accurate picture in small counties than the current methodology that aggregates data across several years.

Thus, we argue that Facebook can serve an intermediary role in augmenting sparse data at a community level. We have shown that it can do so already, but additional health survey data, especially in less extensively measured regions (eg, rural), could only help. Although complete measurement is unfeasible and would render the Facebook modeling moot, ensuring that communities of all types are represented in sufficient number when estimating the model is a necessary step in avoiding the risk of systematic error in its predictions.

The ultimate goal of our analysis of Facebook likes is to establish the potential contribution of big data to research that directly affects government spending and public policy, and—most importantly—contributes to improved population health. At a fraction of the cost of traditional research, data that might seem on its face to have little to do with health can predict epidemic-level health problems such as diabetes and obesity. With the need to augment traditional public health surveillance systems with readily available, cost-effective, and geographically relevant health data, the use of “big epidemiologic data” comes at just the right time.

The nature of the Facebook data source prevents it from being a useful tool in several situations. In the case of very small counties (approximately 9% of the total) and in smaller geographic areas, rounding error becomes so great that estimates cannot be reliably used, even though they may be provided by Facebook. Additionally, Facebook profiles are untested as a tool for tracking the prevalence of infectious diseases. They may be better suited to predicting endemic and ongoing conditions that are unlikely to fluctuate over the course of short time periods.

Further, some might find it counterintuitive that Facebook data are being used to “predict” health data that not only predates it, but to which it is not causally related through any theoretical mechanism. Likes data for a given geographic area should be viewed as a product of sociobehavioral conditions within that region in the same manner that health outcomes are. As such, the likes data can be viewed as an instrument for those conditions, which are causally linked. Although the temporal concerns are not ideal, they are not especially problematic because those health metrics used in this research are not especially prone to fluctuation over short time periods.

Finally, without a clear insight into the manner in which the categories of Facebook likes are constructed and by which individuals are tagged as being interested in a given category, it is difficult to achieve more nuanced insights into the relationships between social network behavior and health outcomes. Unless Facebook becomes more transparent regarding the ways in which these data are compiled, they will remain a “black box” and we must take on faith that the interests and activities being measured are indeed those it claims to measure.

The relationships examined here demonstrate that social media may hold promise to be used as an indicator of local conditions, even those that have little relationship to the activity that takes place on Facebook. As we predicted, significant relationships that extend beyond the predictive power of local demographics exist between an area’s aggregate Facebook behavior and the incidence of diseases and of adverse lifestyle behaviors that very well may lead to those diseases.

We have also indicated the severe shortage of health data that are available in most American counties. Although Facebook data may not reach into every corner of the United States, it seems an effective enough tool to augment the existing county-level data in the majority of counties. With demand for local health data growing, such tools seem far more cost-effective than an increase in survey surveillance regardless of the mode through which it might be conducted.

Whether this data ultimately comes from Facebook or not is of little importance. The online landscape may change and it may provide a different source of data that proves more viable in the future. So long as the source reflects people’s activities in daily life, the same relationships may hold. Even if Facebook does prove to endure as a social institution, however, there is still room for a great deal of improvement on the models presented here. With cooperation from the social media outlets themselves, we may be able to obtain better estimates in categories that align better with our needs. In the end, our data may not suffer because of the rising costs of research. Instead, exploring newly opened avenues of data collection online could lead to more reliable, timely, and cost-effective county-level data than that obtainable from traditional public health surveillance systems as well as serve as an adjunct to those systems.

Acknowledgments

We thank the state BRFSS coordinators for their help in collecting the data used in this analysis and members of the Population Health Surveillance Branch for their assistance in developing the database. The authors would also like to express their thanks to Youjie Huang, MD, MPH, DrPH, Florida Acting BRFSS coordinator, for the data used in the 2010 county-level analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Facebook category structure.

[PDF File (Adobe PDF File), 56KB - jmir_v17i4e98_app1.pdf]

Multimedia Appendix 2

Demographic variable descriptions.

[PDF File (Adobe PDF File), 70KB - jmir_v17i4e98_app2.pdf]

Multimedia Appendix 3

Scree plot for principal components analysis.

[PDF File (Adobe PDF File), 47KB - jmir_v17i4e98_app3.pdf]

Multimedia Appendix 4

Rotated (orthogonal varimax) factors.

[PDF File (Adobe PDF File), 50KB - jmir_v17i4e98_app4.pdf]

Multimedia Appendix 5

Factor loadings.

[PDF File (Adobe PDF File), 54KB - jmir_v17i4e98_app5.pdf]

References

1. Hand E. Citizen science: People power. Nature 2010 Aug 5;466(7307):685-687. [doi: 10.1038/466685a] [Medline: 20686547]
2. Brownstein CA, Brownstein JS, Williams DS, Wicks P, Heywood JA. The power of social networking in medicine. Nat Biotechnol 2009 Oct;27(10):888-890. [doi: 10.1038/nbt1009-888] [Medline: 19816437]
3. Boicey C. Innovations in social media: the MappyHealth experience. Nursing Management 2013 Mar;44(3):10-11. [doi: 10.1097/01.NUMA.0000427191.36468.65] [Medline: 23435102]
4. 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]
5. Chawla NV, Davis DA. Bringing big data to personalized healthcare: a patient-centered framework. J Gen Intern Med 2013 Sep;28 Suppl 3:S660-S665 [FREE Full text] [doi: 10.1007/s11606-013-2455-8] [Medline: 23797912]
6. Rogstadius J, Vukovic M, Teixeira C, Kostakos V, Karapanos E, Laredo J. CrisisTracker: Crowdsourced social media curation for disaster awareness. IBM J Res & Dev 2013 Sep;57(5):4:1-4:13. [doi: 10.1147/JRD.2013.2260692]
7. 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]
8. Weitzman ER, Kelemen S, Mandl KD. Surveillance of an online social network to assess population-level diabetes health status and healthcare quality. Online J Public Health Inform 2011;3(3):1-1 [FREE Full text] [doi: 10.5210/ojphi.v3i3.3797] [Medline: 23569613]
9. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
10. Salathe M, Bengtsson L, Bodnar TJ. Digital epidemiology. PLoS Comput Biol 2012;8(7):1-5 [FREE Full text] [doi: 10.1371/journal.pcbi.1002616]
11. Morse SS. Public health surveillance and infectious disease detection. Biosecur Bioterror 2012 Mar;10(1):6-16. [doi: 10.1089/bsp.2011.0088] [Medline: 22455675]
12. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg 2012 Jan;86(1):39-45 [FREE Full text] [doi: 10.4269/ajtmh.2012.11-0597] [Medline: 22232449]
13. Chunara R, Bouton L, Ayers JW, Brownstein JS. Assessing the online social environment for surveillance of obesity prevalence. PLoS One 2013;8(4):e61373 [FREE Full text] [doi: 10.1371/journal.pone.0061373] [Medline: 23637820]
14. Ayers JW, Althouse BM, Allem JP, Childers MA, Zafar W, Latkin C, et al. Novel surveillance of psychological distress during the great recession. J Affect Disord 2012 Dec 15;142(1-3):323-330. [doi: 10.1016/j.jad.2012.05.005] [Medline: 22835843]
15. Schmidt CW. Trending now: using social media to predict and track disease outbreaks. Environ Health Perspect 2012 Jan;120(1):30-33.

16. Waggoner MR. Parsing the peanut panic: the social life of a contested food allergy epidemic. Soc Sci Med 2013 Aug;90:49-55 [FREE Full text] [doi: 10.1016/j.socscimed.2013.04.031] [Medline: 23746608]
17. Chary M, Genes N, McKenzie A, Manini AF. Leveraging social networks for toxicovigilance. J Med Toxicol 2013 Jun;9(2):184-191 [FREE Full text] [doi: 10.1007/s13181-013-0299-6] [Medline: 23619711]
18. Lauer MS. Time for a creative transformation of epidemiology in the United States. JAMA 2012 Nov 7;308(17):1804-1805. [doi: 10.1001/jama.2012.14838] [Medline: 23117782]
19. Khoury MJ, Lam TK, Ioannidis JPA, Hartge P, Spitz MR, Buring JE, et al. Transforming epidemiology for 21st century medicine and public health. Cancer Epidemiol Biomarkers Prev 2013 Apr;22(4):508-516 [FREE Full text] [doi: 10.1158/1055-9965.EPI-13-0146] [Medline: 23462917]
20. Crawford CAG, Okoro CA, Akcin HM, Dhingra S. An experimental study using opt-in Internet panel surveys for behavioral health surveillance. Online J Public Health Inform 2012;5(1):e24-e24.
21. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. J Clin Epidemiol 2010 Nov;63(11):1169-1178 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.11.021] [Medline: 20688473]
22. Minniear TD, McIntosh EB, Alexander N, Weidle PJ, Fulton J. Using electronic surveys to gather information on physician practices during a response to a local epidemic-Rhode Island. Annual Epidemiol 2011;2013:1-3.
23. 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]
24. Lyon A, Nunn M, Grossel G, Burgman M. Comparison of web-based biosecurity intelligence systems: BioCaster, EpiSPIDER and HealthMap. Transbound Emerg Dis 2012 Jun;59(3):223-232. [doi: 10.1111/j.1865-1682.2011.01258.x] [Medline: 22182229]
25. Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. PLoS Medicine 2013;10(4):1-4 [FREE Full text]
26. 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]
27. Yoon S, Elhadad N, Bakken S. A practical approach for content mining of Tweets. Am J Prev Med 2013 Jul;45(1):122-129 [FREE Full text] [doi: 10.1016/j.amepre.2013.02.025] [Medline: 23790998]
28. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: A literature review and analysis of affordances. J Biomed Inform 2013 May:957-969.
29. Facebook. 2013. URL: https://www.facebook.com/unsupportedbrowser [accessed 2015-03-22] [WebCite Cache ID 6XEQbNrA1]
30. Twitter. URL: https://twitter.com/ [accessed 2015-03-22] [WebCite Cache ID 6XEQffz6d]
31. Instagram. URL: https://instagram.com/ [accessed 2015-03-22] [WebCite Cache ID 6XEQjzEYH]
32. tumblr. URL: https://www.tumblr.com/ [accessed 2015-03-22] [WebCite Cache ID 6XET0yDHr]
33. Google. URL: http://www.google.com/ [accessed 2015-03-22] [WebCite Cache ID 6XETDHclP]
34. Amazon. URL: http://www.amazon.com/ [accessed 2015-03-22] [WebCite Cache ID 6XETGzlh0]
35. Murray D, Durrell K. Inferring demographic attributes of anonymous Internet users. In: Masand B, Spiliopoulou M, editors. Web Usage Analysis and Demographic Profiling. Berlin: Springer; 2000:7-20.
36. Goel S, Hofman JM, Sirer MI. Who does what on the Web: A large-scale study of browsing behavior. 2012 Presented at: ICWSM; June 4-7, 2012; Toronto, ON.
37. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci USA 2013 Apr 9;110(15):5802-5805 [FREE Full text] [doi: 10.1073/pnas.1218772110] [Medline: 23479631]
38. Tong ST. Facebook use during relationship termination: uncertainty reduction and surveillance. Cyberpsychol Behav Soc Netw 2013;20:1-6.
39. Mervis J. US science policy: agencies rally to tackle big data. Science 2012 Apr 6;336(6077):22. [doi: 10.1126/science.336.6077.22] [Medline: 22491835]
40. Bond RM, Fariss CJ, Jones JJ, Kramer ADI, Marlow C, Settle JE, et al. A 61-million-person experiment in social influence and political mobilization. Nature 2012 Sep 13;489(7415):295-298 [FREE Full text] [doi: 10.1038/nature11421] [Medline: 22972300]
41. Chang A, Anderson EE, Turner HT, Shoham D, Hou SH, Grams M. Identifying potential kidney donors using social networking web sites. Clin Transplant 2013;27(3):E320-E326 [FREE Full text] [doi: 10.1111/ctr.12122] [Medline: 23600791]
42. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N. Tastes, ties, and time: A new social network datset using Facebook.com. Social Networks 2008;30:330-342.
43. Jernigan C, Mistree BF. Firstmonday.org. 2009 Sep 25. Gaydar: Facebook friendships expose sexual orientation URL: http://firstmonday.org/article/view/2611/2302 [accessed 2015-03-22] [WebCite Cache ID 6XETbt4p6]

44. Timian A, Rupcic S, Kachnowski S, Luisi P. Do patients "like" good care? Measuring hospital quality via Facebook. Am J Med Qual 2013:374-382.
45. Mobley LR, Finkelstein EA, Khavjou OA, Will JC. Spatial analysis of body mass index and smoking behavior among WISEWOMAN participants. J Womens Health (Larchmt) 2004 Jun;13(5):519-528. [Medline: 15266669]
46. Schuit AJ, van Loon AJM, Tijhuis M, Ocké M. Clustering of lifestyle risk factors in a general adult population. Prev Med 2002 Sep;35(3):219-224. [Medline: 12202063]
47. Centers for Disease Control and Prevention. 2015. National Vital Statistics System URL: http://www.cdc.gov/nchs/nvss.htm [accessed 2015-03-22] [WebCite Cache ID 6XETguXAY]
48. Centers for Disease Control and Prevention. 2015. Behavioral Risk Factor Surveillance System URL: http://http://www.cdc.gov/brfss/ [accessed 2015-03-22] [WebCite Cache ID 6XEToDiNo]
49. Centers for Disease Control and Prevention. 2013. SMART: BRFSS City and County Data and Documentation URL: http://www.cdc.gov/brfss/smart/smart_data.htm [accessed 2015-03-22] [WebCite Cache ID 6XEVhKu5Y]
50. Facebook. Facebook ads API URL: https://www.facebook.com/unsupportedbrowser [accessed 2015-03-22] [WebCite Cache ID 6XEVjDSBE]
51. Adler NE, Boyce WT, Chesney MA, Folkman S, Syme SL. Socioeconomic inequalities in health. No easy solution. JAMA 1993;269(24):3140-3145. [Medline: 8505817]
52. Murray CJ, Abraham J, Ali MK. The state of US Health, 1990-2010: burden of diseases, injuries, and risk factors. JAMA 2013 Jul;310(6):591-608.
53. United States' Census 2010. 2010 Census URL: http://www.census.gov/2010census/ [accessed 2015-03-22] [WebCite Cache ID 6XE19WZ2R]
54. United States Census Bureau. USA counties information URL: http://www.census.gov/support/USACdata.html [accessed 2015-03-22] [WebCite Cache ID 6XE1BXbjH]
55. Efron B, Tibshirani R. An Introduction to the Bootstrap. London: CRC press; 1993.
56. Cattell RB. The scree test for the number of factors. Multivariate Behav Res 1966;1(2):245-276.

Abbreviations

API: application program interface
BRFSS: Behavioral Risk Factor Surveillance System
NVSS: National Vital Statistics System
OLS: ordinary least squares
SES: socioeconomic status
SMART: Selected Metropolitan/Micropolitan Area Risk Trends

Edited by G Eysenbach; submitted 03.11.14; peer-reviewed by S Goel, G Khalil, R Bright; comments to author 24.11.14; revised version received 23.02.15; accepted 02.03.15; published 20.04.15.

Please cite as:
Gittelman S, Lange V, Gotway Crawford CA, Okoro CA, Lieb E, Dhingra SS, Trimarchi E
A New Source of Data for Public Health Surveillance: Facebook Likes
J Med Internet Res 2015;17(4):e98
URL: http://www.jmir.org/2015/4/e98/
doi:10.2196/jmir.3970
PMID:25895907

Original Paper

The Effect of Modality and Narration Style on Recall of Online Health Information: Results From a Web-Based Experiment

Nadine Bol[1]; Julia CM van Weert[1], PhD; Hanneke CJM de Haes[2], PhD; Eugene F Loos[1], PhD; Ellen MA Smets[2], PhD

[1]Amsterdam School of Communication Research / ASCoR, Department of Communication Science, University of Amsterdam, Amsterdam, Netherlands
[2]Department of Medical Psychology, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Nadine Bol
Amsterdam School of Communication Research / ASCoR
Department of Communication Science
University of Amsterdam
Nieuwe Achtergracht 166
Amsterdam, 1018 WV
Netherlands
Phone: 31 205252113
Fax: 31 205253681
Email: n.bol@uva.nl

Abstract

Background: Older adults are increasingly using the Internet for health information; however, they are often not able to correctly recall Web-based information (eHealth information). Recall of information is crucial for optimal health outcomes, such as adequate disease management and adherence to medical regimes. Combining effective message strategies may help to improve recall of eHealth information among older adults. Presenting information in an audiovisual format using conversational narration style is expected to optimize recall of information compared to other combinations of modality and narration style.

Objective: The aim of this paper is to investigate the effect of modality and narration style on recall of health information, and whether there are differences between younger and older adults.

Methods: We conducted a Web-based experiment using a 2 (modality: written vs audiovisual information) by 2 (narration style: formal vs conversational style) between-subjects design (N=440). Age was assessed in the questionnaire and included as a factor: younger (<65 years) versus older (≥65 years) age. Participants were randomly assigned to one of four experimental webpages where information about lung cancer treatment was presented. A Web-based questionnaire assessed recall of eHealth information.

Results: Audiovisual modality (vs written modality) was found to increase recall of information in both younger and older adults (*P*=.04). Although conversational narration style (vs formal narration style) did not increase recall of information (*P*=.17), a synergistic effect between modality and narration style was revealed: combining audiovisual information with conversational style outperformed combining written information with formal style (*P*=.01), as well as written information with conversational style (*P*=.045). This finding suggests that conversational style especially increases recall of information when presented audiovisually. This combination of modality and narration style improved recall of information among both younger and older adults.

Conclusions: We conclude that combining audiovisual information with conversational style is the best way to present eHealth information to younger and older adults. Even though older adults did not proportionally recall more when audiovisual information was combined with conversational style than younger adults, this study reveals interesting implications for improving eHealth information that is effective for both younger and older adults.

(J Med Internet Res 2015;17(4):e104) doi:10.2196/jmir.4164

KEYWORDS

instructional films and videos; narration; personal narratives; age groups; consumer health information; memory; mental recall; patient education

Introduction

Background

Older adults have been one of the fastest growing groups using the Internet [1]. Recent figures show that the number of older adults using the Internet in Western countries, such as the United States and the Netherlands, has nearly doubled over the past few years [2,3]. More than half of these older adults used the Internet for accessing health information [3,4]. The Internet is seen as a relevant source of gathering health information among older adults [5] and has helped many older adults to fulfill a wide range of information and support needs [6,7]. However, simply having access to online health information (subsequently called eHealth information in this paper) does not necessarily mean that people find and understand such information. eHealth literacy, the ability to seek, find, understand, and act on health information from electronic sources to solve a health problem [8], is considered lower among older adults [9]. This may influence the extent to which health information is recalled. Recall of information is the ability to reproduce information and is crucial for optimal health outcomes, such as adequate disease management [10] and adherence to medical regimes [11,12].

Thus, effective message strategies need to be found to communicate eHealth information in such a way that older adults are able to recall information. However, previous research focusing on effective ways to present eHealth information to older adults in particular is lacking. The cognitive theory of multimedia learning (CTML) presents strategies that might enhance recall in general and could be applied to how older adults learn from eHealth materials. One strategy is to use audiovisual materials. The idea that people learn information more deeply from audiovisual materials compared to written materials is called the *modality effect* [13]. Another strategy described by the CTML is to use conversational narration style in messages. People are expected to learn information better when presented in conversational narration style rather than formal narration style, a phenomenon referred to as the *personalization effect* [14]. Surprisingly, to the best of our knowledge, no studies to date have examined the effects of combining modality (written vs audiovisual) and narration style (formal vs conversational) in eHealth messages on information recall. Combining such strategies might result in synergistic effects, which occur when the combined effect of these strategies exceeds the sum of their individual effects [15]. Furthermore, although previous research has shown support for modality and narration style strategies in younger adults [16,17], empirical evidence on the effects in older adults is particularly lacking.

The aim of our study is to identify effective ways of presenting Web-based information to enhance recall of eHealth information among older adults in particular. We test the (synergistic) effect of modality and narration style, and whether there are differences between younger and older adults. In assessing age differences, we chose participants younger than 65 years for the younger age group and 65 years and older as the older age group. These two age groups are generally considered suitable for separate analysis in aging research [18,19].

Modality: Written Versus Audiovisual Information

The CTML attempts to explain how people learn and recall information in multimedia environments [20]. It is based on the assumption that people have separate information processing systems for visual and auditory materials [21] and that people are limited in the amount of material they can process in each of these systems at one time [22]. These assumptions are derived from the dual coding theory [21,22] and cognitive load theory [13]. The modality principle describes that one’s limited working memory capacity can be effectively expanded by addressing one or more sensory modalities, such as audiovisual information, instead of written information only [23]. A meta-analysis indeed showed an overall benefit in recall of information when information was presented in an audiovisual format compared to written format [16]. However, this meta-analysis did not include studies that were conducted in the field of eHealth information. We could nevertheless expect that audiovisual information would enhance recall of eHealth information as well. We therefore suggest H1a: Exposure to audiovisual information, compared to written information, has a positive effect on recall of eHealth information.

Whereas the modality effect among younger adults is repeatedly found in research [16], fewer studies have focused specifically on its effect among older adults [24]. The cognitive aging principle in multimedia learning [25] provides an integrative framework where the cognitive load theory is combined with general views of cognitive aging. This theory states that older adults’ limited working memory might be expanded by multimodal presentation of information. Older adults usually have a smaller “total cognitive capacity” than younger adults and might therefore benefit more from multimodal information than younger adults [26]. Previous research has demonstrated a benefit of multimedia messages for older adults by reducing cognitive load and decreasing learning time of information [24]. Even though this study did not present eHealth information to older adults, audiovisual information might be as effective for older adults when used in a Web-based context. We might therefore expect that audiovisual information positively influences older adults’ memory for eHealth information. Therefore, our second hypothesis H1b states: The effects of exposure to audiovisual information, compared to written information, on recall of eHealth information are greater in older adults than in younger adults.

Narration Style: Formal Versus Conversational Style

Conversational style is often used in narrative communication, such as testimonials and personal stories or a description of an individual experience [27]. By presenting information from a first-person perspective using a more informal approach, it feels like the person in the message is talking to you [28]. However, health information has been traditionally presented in a formal style, presenting factual information, such as expressing professional opinions [29]. Unfortunately, a substantial number of people do not understand such formal health messages well enough to make informed decisions and act accordingly [30]. In an attempt to simplify health information, using a conversational style in narrative communication has shown to serve as an effective way of educating people about health

[27,31]. Conversational style is more likely to be recalled because of conversational rules, such as commitment to try to understand a narrator's story [32]. Moreover, conversational style has been found to increase recall of information among a variety of learner profiles, that is, different levels of education [33]. For this reason, we hypothesize H2a: Exposure to information in a conversational style, compared to information in a formal style, has a positive effect on recall of eHealth information.

In addition to the hypothesized effect of conversational style among both younger and older adults, it is expected that older adults may benefit in particular from having eHealth information presented in conversational style. Older adults tend to have better narrative recall and thus remember stories in more accurate detail than younger adults [34]. Using conversational style in eHealth information might therefore especially improve recall of information in older adults. Therefore, we expect H2b: The effects of exposure to information in a conversational style, compared to information in a formal style, on recall of eHealth information are greater in older adults than in younger adults.

Synergistic Effects: Combining Modality and Narration Style

To our knowledge, no research to date has studied the synergistic effects of combining modality and narration style in enhancing recall of information. Nevertheless, various message strategies and styles are often combined in one health message, and it is likely that these message strategies have individual as well as synergistic effects [35]. For instance, one study found that tailoring Web-based information to individual characteristics was effective in realizing smoking cessation, however, only when information was presented in audiovisual format and not in written format [36]. This finding indicates that combining several message characteristics might optimize the effectiveness of a message. As we hypothesized that audiovisual information will outperform written information and conversational narration style will outperform formal narration style, we might expect H3a: Combining audiovisual information with conversational style, compared to other combinations of modality and narration style, has a positive effect on recall of eHealth information.

Moreover, as we hypothesized that especially older adults will benefit from audiovisual information and conversational narration style, we might also expect that combining these two might especially enhance older adults' recall of information. Our final hypothesis H3b therefore states: The effects of combining audiovisual information with conversational style, compared to other combinations of modality and narration style, on recall of eHealth information are greater in older adults than in younger adults.

Methods

Design

For this experiment, a 2 (modality: written vs audiovisual information) by 2 (narration style: formal vs conversational style) between-subjects design was used. Age was assessed in the questionnaire and included as a factor: younger (<65 years) versus older (≥65 years) age. We used a Netherlands Cancer Institute (NKI) webpage where information about radio frequency ablation (RFA) treatment was given. RFA is a minimally invasive treatment, using radio frequency waves to destroy lung tumors.

Ethical approval for this study was granted by the Institutional Review Board of the Amsterdam School of Communication Research (reference number 2012-CW-33).

Participants

Participants from the general population were selected from a large respondent pool by the ISO certified market research company PanelClix to participate in a Web-based survey. This Web-based setting allowed participants to engage in the experiment from their home computer, creating a natural and realistic setting to evaluate the stimulus materials. PanelClix made a random selection of their panel members to participate in the study stratified by age (<65 years vs ≥65 years). To be eligible to take part in the Web-based survey, participants had to be older than 18 years, able to read and write in Dutch, and should have no prior knowledge on RFA treatment, because this could influence the recall scores. An invitation for the study was available for PanelClix members on the PanelClix website. A total of 796 unique participants entered the survey, of which 788 unique participants started the survey (participation rate=99.0%), and 490 unique participants completed the Web-based survey (completion rate=62.2%). Duplicate entries were avoided by assigning pid-codes to participants, which were unique codes assigned to participants by PanelClix. Checking these pid-codes revealed 4 participants who completed the survey twice. In these cases, the second entry was excluded from the dataset (n=4). Participants who had prior knowledge on RFA treatment, that is, scoring higher than 4 on a 7-point Likert scale (n=23), were excluded as well. Since usability testing of the questionnaire showed that at least 10 minutes were needed to complete the survey, participants were excluded when they completed the survey in less than 10 minutes (n=15). Furthermore, participants were excluded when they viewed the experimental stimulus for less than 4 seconds (n=1), had not been exposed to the experimental stimulus due to technical issues (n=5), had used another source than the experimental stimulus to answer the recall questions (n=1), or reported to suffer from short-term memory loss (n=1). This resulted in a total of 440 participants included in the dataset.

Procedure

When participants entered the Web-based survey, they were informed about the survey length, participants' rights (eg, anonymity), and the contact details of the research institute and principle investigator. On the next page, participants were exposed to questions about gender, age, educational level, prior medical knowledge about lung cancer and RFA treatment, and Internet use. Based on age, participants were assigned to the younger or older age strata, and were randomly assigned to one of four experimental conditions. They were instructed to pay careful attention to the RFA information and could look at the information as long as they liked. Participants were not able to return to the written or audiovisual information (depending on condition) after continuing to the next page. The written information versions were presented on a webpage and the audiovisual information versions through a Web-based video. In the audiovisual conditions, participants were instructed to turn on their speakers. After exposure to the experimental stimulus, recall questions were shown. Upon completion, participants received credit points from PanelClix.

Materials

Using the existing text (formal written condition, 245 words) of the NKI website, the conversational text was written (330 words). The formal audiovisual information was identical to the formal written information, and the conversational audiovisual information was identical to the conversational written information. The conversational versions were personalized in two ways: first, we changed the third-person perspective into a first-person perspective as if the narrator was talking to you [28] (eg, “A special needle guided by a CT scanner will be inserted into the cancerous lung tumor” vs “A special needle guided by a CT scanner was inserted into my cancerous lung tumor”). Second, we added sentences about personal experience to make the story more conversational as well (eg, “[…], depending on the location of the tumor” vs “I myself received sedation, but that depended on the location of the tumor”). In the formal audiovisual conditions, a doctor was videotaped behind a desk and in the conversational audiovisual conditions, a patient was filmed sitting on a couch. The same actors starred as the doctor and patient to ensure that narration style effects could be attributed to changes in narration style rather than changes in narrator. In addition, both the formal and conversational audiovisual information was filmed twice to create a formal and conversational video starring a younger narrator as well as a formal and conversational video starring an older narrator. This resulted in four videos in which information was presented by (1) a younger doctor (2:01 min; see Multimedia Appendix 1), (2) an older doctor (2:05 min; see Multimedia Appendix 2), (3) a younger patient (2:53 min; see Multimedia Appendix 3), or (4) an older patient (2:34 min; see Multimedia Appendix 4). This was done to control for potential identification effects with a specific narrator because of age similarity. However, there were no differences in identifying with the younger and older actor ($F_{1,245}$=0.06, P=.86, η_p^2=0.00), regardless of participants’ own age ($F_{1,245}$=0.13, P=.72, η_p^2=0.00). Therefore, this additional design factor was not considered while analyzing the data. Thus, four experimental conditions were analyzed: (1) written information in formal style, (2) written information in conversational style, (3) audiovisual information in formal style, and (4) audiovisual information in conversational style. Compared to previous research on the effects of modality and narration style, our materials were similar in terms of length [16,17].

Measures

Background Characteristics

Background measures included participants’ age, gender, level of education, Internet use, and prior medical knowledge. Education was divided into low level of education (primary education, lower vocational education, preparatory secondary vocational education, and intermediate secondary vocational education), middle level of education (senior secondary vocational education and university preparatory vocational education), and high level of education (higher vocational education and university). Internet use was assessed through average number of hours spent per week on the Internet. Prior medical knowledge about lung cancer in general and RFA knowledge specifically was measured by two items asking how much knowledge participants perceived to have about lung cancer and RFA treatment respectively using a 7-point scale (1=no knowledge at all, 7=very much knowledge).

Recall of eHealth Information

Information recall was measured using the Netherlands Patient Information Recall Questionnaire [37]. Questions were created from the RFA information and were pretested among 12 students. This resulted in 11 open-ended recall questions, such as “Could you please name the most common complication during an RFA treatment?” All questions were accompanied by a textbox for participants to provide their answer. Recall scores could range from 0 (not recalled), to 1 (recalled partially), to 2 (recalled correctly). A codebook was used for allocating scores to each of the recall questions. Two independent coders used this codebook and double coded 14.1 % (62/440) of the recall scores. Interrater reliability appeared to be good (mean kappa=.84, range 0.65-1.00). The 11 items were computed into a total recall score, ranging from 0-22. Additionally, percentages of the recall scores were calculated by dividing the participant’s total score by 22.

Statistical Analysis

For testing successful randomization, analyses of variance and chi-square tests were conducted to check for unequal distribution of background variables across the 2 (modality) × 2 (narration style) × 2 (age groups) experimental design. Educational level and medical knowledge about lung cancer were found to be unequally distributed (respectively χ^2_{14}=24.72, P=.04 and $F_{7,432}$=2.06, P=.047, η_p^2=0.03). Moreover, previous studies have identified educational level and prior topic knowledge as important predictors of information recall [38]. We therefore included educational level and medical knowledge about lung cancer as covariates in all analyses. For testing the main effects of modality and narration style, an ANCOVA (analysis of covariance) was conducted with recall of information as dependent variable, modality, narration style, and age groups as factors, and education and medical knowledge as covariates. The synergistic effects between modality and narration style were tested in an ANCOVA using one variable for modality and narration to measure the simple contrast effects between conversational audiovisual information and the three other combinations of modality and narration style. To test whether the effect of modality and narration style was greater among older adults than younger adults, the data file was split on age group and the above described analyses were repeated.

Results

Sample Characteristics

Of the participants who filled out the Web-based questionnaire (53.9%, 237/440 male and 46.1%, 203/440 female), 53.6% were younger (236/440; <65 years; mean 41.78, SD 12.69) and 46.4% were older (204/440; ≥65 years; mean 69.44, SD 4.13). Besides age ($F_{1,438}$=887.88, P<.001, η_p^2=0.67), the two age groups significantly differed in educational level (χ^2_2=18.61, P<.001; see Table 1 for sample characteristics).

Table 1. Sample characteristics (N=440).[a]

	Younger adults (<65 years) (n=236)	Older adults (≥65 years) (n=204)
Gender, n (%)		
Male	123 (52.1)	114 (55.9)
Female	113 (47.9)	90 (44.1)
Age in years, mean (SD) range	41.78 (12.69) 18-64	69.44 (4.13)[b]65-85
Education level, n (%)		
Low	50 (21.2)	81 (40.1)[b]
Middle	104 (44.1)	69 (34.2)[c]
High	82 (34.7)	52 (25.7)
Internet use (hours per week), mean (SD) range	18.21 (11.72) 1-60	17.83 (11.59) 1-60
Medical knowledge about lung cancer[d], mean (SD) range	2.88 (1.41) 1-6	2.66 (1.37) 1-7
Medical knowledge about RFA[d], mean (SD) range	1.71 (1.06) 1-4	1.54 (0.94) 1-4

[a] Not all figures add up to 100% due to missing data. Conditions stratified by age only significantly differed on education.

[b] Differs significantly from younger adults (P<.001).

[c] Differs significantly from younger adults (P=.04).

[d] A higher score indicates more knowledge, ranging from 1 to 7. RFA: Radio Frequency Ablation.

Effects of Modality on Recall of eHealth Information

The first hypothesis predicted a positive effect of audiovisual information (vs written information) on recall of eHealth information (H1a), which was greater for older adults in particular (H1b). Modality significantly influenced recall of health information ($F_{1,413}$=4.48, P=.04, η_p^2=0.01). As hypothesized, the audiovisual conditions (mean 7.60, SE 0.31, 95% CI 7.00-8.21) resulted in significantly higher recall scores than the written conditions (mean 6.55, SE 0.39, 95% CI 5.79-7.32). However, older adults did not recall more from audiovisual information compared to written information ($F_{1,189}$=1.59, P=.21, η_p^2=0.01) than younger adults ($F_{1,223}$=3.09, P=.08, η_p^2=0.01). These results, thus, partially confirm our first hypothesis. Recall scores across modality and age groups appear in Table 2.

Table 2. Main effects of modality on recall of eHealth information in younger and older adults.[a]

		Recall of health information		
Group	n[b]	% recall	Mean (SE)	95% CI
All participants				
Written information	165	29.8	6.55 (0.39)	5.79-7.32
Audiovisual information	273	34.5	7.60[c](0.31)	7.00-8.21
Younger participants				
Written information	85	28.2	6.20 (0.55)	5.12-7.27
Audiovisual information	151	34.1	7.50 (0.42)	6.68-8.32
Older participants				
Written information	80	31.4	6.91 (0.55)	5.83-7.99
Audiovisual information	122	35.0	7.70 (0.45)	6.82-8.59

[a] Adjusted for education level and medical knowledge about lung cancer. Recall of information ranges from 0-22. Percentage of correct recall is based on mean scores divided by 22. The higher the score, the more information was recalled correctly.

[b] The category sizes differ because of 2 fewer cases due to missing covariate values.

[c] Differs significantly compared to written information (P=.04).

Effects of Narration Style on Recall of eHealth Information

Our second hypothesis predicted a positive effect of conversational narration style (vs formal narration style) on recall of eHealth information (H2a), which was expected to be greater for older adults in particular (H2b). However, no main effect of narration style on recall of information was found ($F_{1,413}$=1.89, P=.17, η_p^2=0.01). The means in Table 3 show that using conversational style does not significantly increase recall of information (mean 7.42, SE 0.35, 95% CI 6.74-8.10) compared to formal style (mean 6.74, SE 0.35, 95% CI 6.04-7.43). Furthermore, older adults did not recall more information when conversational style (vs formal style) was used ($F_{1,189}$=1.78, P=.18, η_p^2=0.01) compared to younger adults ($F_{1,223}$=0.51, P=.48, η_p^2=0.00). This was not expected and, therefore, our second hypothesis was rejected. Recall scores across narration style and age groups are shown in Table 3.

Table 3. Main effects of narration style on recall of eHealth information in younger and older adults.[a]

		Recall of health information		
Group	n[b]	% recall	Mean (SE)	95% CI
All participants				
Formal style	214	30.6	6.74 (0.35)	6.04-7.43
Conversational style	224	33.7	7.42 (0.35)	6.74-8.10
Younger participants				
Formal style	119	30.0	6.61 (0.48)	5.67-7.55
Conversational style	117	32.2	7.09 (0.50)	6.11-8.06
Older participants				
Formal style	95	31.2	6.86 (0.52)	5.84-7.88
Conversational style	107	35.2	7.75 (0.48)	6.80-8.70

[a] Adjusted for education level and medical knowledge about lung cancer. Recall of information ranges from 0-22. Percentage of correct recall is based on mean scores divided by 22. The higher the score, the more information was recalled correctly.

[b] The category sizes differ because of 2 fewer cases due to missing covariate values.

Synergistic Effects

Our third hypothesis concerned the synergistic effect between modality and narration style (H3a), which was also expected to be greater for older adults (H3b). We expected that combining audiovisual information with conversational style would outperform other combinations of modality and narration style. In support of our hypothesis, we found that combining audiovisual information with conversational style resulted in the most favorable recall scores compared to combining written information with formal style (contrast estimate=-1.73, SE 0.70, P=.01, 95% CI -3.11 to -0.35]), as well as compared to combining written information with conversational style (contrast estimate=-1.40, SE 0.70, P=.045, 95% CI -2.76 to -0.03). This finding suggests that conversational style especially increases recall of information when presented as audiovisual information, but not when presented as written information (see Figure 1).

The expected synergistic combination of audiovisual information and conversational style (eg, vs combining written information with formal style) did not particularly improve recall of information among older adults (contrast estimate=-1.73, SE 0.95, P=.07, 95% CI -3.60 to 0.15) compared to younger adults (contrast estimate=-1.80, SE 1.03, P=.08, 95% CI -3.84 to 0.24). Our data therefore partially confirm our third hypothesis. The synergistic effects between modality and narration style on recall of information (for all participants and stratified by age group) are provided in Table 4.

Figure 1. Combining audiovisual information with conversational narration style results in highest recall of eHealth information among younger and older adults. The bars and error bars represent the mean recall scores and 95% confidence intervals respectively.

Table 4. Synergistic effects of combining modality and narration style on recall of eHealth information in younger and older adults.[a]

Group	n[b]	Recall of health information % recall	Mean (SE)	95% CI
All participants				
Written				
Formal	82	29.1	6.39 (0.56)	5.29-7.48
Conversational	83	30.6	6.72 (0.54)	5.65-7.79
Audiovisual				
Formal	132	32.2	7.09 (0.44)	6.23-7.94
Conversational	141	36.9	8.12[c,d](0.43)	7.26-8.97
Younger participants				
Written				
Formal	44	28.0	6.16 (0.77)	4.66-7.67
Conversational	41	28.3	6.23 (0.78)	4.69-7.77
Audiovisual				
Formal	75	32.1	7.06 (0.57)	5.94-8.18
Conversational	76	36.1	7.94 (0.61)	6.73-9.15
Older participants				
Written				
Formal	38	30.1	6.61 (0.81)	5.03-8.19
Conversational	42	32.8	7.21 (0.75)	5.73-8.68
Audiovisual				
Formal	57	32.4	7.12 (0.66)	5.83-8.41
Conversational	65	37.7	8.29 (0.61)	7.09-9.49

[a] Adjusted for education level and medical knowledge about lung cancer. Recall of information ranges from 0-22. Percentage of correct recall is based on mean scores divided by 22. The higher the score, the more information was recalled correctly.

[b] The category sizes differ because of 2 fewer cases due to missing covariate values.

[c] Differs significantly compared to formal written information (*P*=.01).

[d] Differs significantly compared to conversational written information (*P*=.045).

Discussion

Principal Findings

In this study, we aimed to identify effective ways of presenting eHealth information to enhance recall of information among older adults in particular. We examined the (synergistic) effect of modality and narration style on recall of health information, and whether there are differences between younger and older adults. Our results support the modality effect as proposed by the CTML [13]. Younger and older individuals who were exposed to audiovisual information recalled more health information than those who were exposed to written information. Our findings do not show support for the personalization effect, another principle proposed by the CTML [14]. Health information was not better recalled when presented in conversational style than in formal style. However, combining audiovisual information with conversational style led to the highest recall scores, outperforming combining written information with formal style and written information with conversational style. This underscores that conversational style is especially effective in improving recall of health information when presented audiovisually, rather than when presented as written information.

Even though we found that older adults benefited from audiovisual information with respect to better recall of information, they did not proportionally benefit more than younger adults, as proposed in the cognitive aging principle in multimedia learning [25]. This might be explained by the fact that audiovisual information is not self-paced, whereas written information is. Previous research has indicated that self-pacing of information plays an important role in older adults' recall of information [39]. When older adults have the option to self-pace information, they are able to take the time they need to process information, which may result in information recall that is comparable to that of younger adults (personal communication by Bol, Van Weert, Loos, Romano Bergstrom, Bolle & Smets, 2014). Likewise, another study showed that older adults need more time than younger adults to recall equivalent amounts of information [40]. Hence, older adults might benefit most from self-paced information. However, audiovisual information is traditionally not self-paced. A recent experimental study, in which self-paced written information was compared with self-paced spoken information, revealed that self-paced spoken information outperformed self-paced written information in older adults with limited health literacy [41]. When exposed to self-paced spoken animations (ie, spoken information with simulated motion pictures), older adults with low health literacy recalled the same amount of information as their high health-literate counterparts. Older people with health disparities might therefore especially benefit more from audiovisual information with respect to better recall. As health literacy was not measured, and individuals with limited health literacy might have been underrepresented in the current sample, this seems worth investigating in future research.

This study provides evidence for effective Web-based communication strategies by revealing the promising effects of combining message factors, such as audiovisual information and conversational narration style, to enhance information recall in both younger and older adults. In addition to this practical contribution, the results of this study also add to the current synergy literature. Our results suggest that synergistic effects do not occur only at the level of combining multiple media as suggested by Naik and Raman [15], but also at the level of combining multiple message characteristics, in this case, modality and narration style. Nonetheless, we need to use caution in generalizing these results as there are more message characteristics to explore and effective combinations to discover (for an overview of other message characteristics, see [28]).

Limitations

This study has several limitations. First, the manipulation we used for narration style could be considered as a study limitation. In Web-based written material, conversational style can be clearly manipulated by changing passages in the text [28]. However, when working with audiovisual information, we dealt with visual changes as well, that is, changing the type of narrator (doctor vs patient). Having a doctor versus a patient explain information changes source attributes, which might have changed the perceived source expertise and the perceived level of authority [42]. This change might have confounded the association between conversational style and recall of information in the audiovisual conditions, providing an alternative explanation for the synergistic effect found when combining audiovisual information with conversational style. Furthermore, it should be noted that simply putting written information into spoken format will not fully ensure equivalence. Features that determine the listenability of spoken messages are not the same as the features that determine the readability of written messages [43].

Second, although combining audiovisual information with conversational narration style led to the highest recall scores, it is important to bear in mind that the overall recall scores in this study were low. It has been estimated that 40%-80% of medical information is immediately forgotten [44], which has also been found with regard to recall of Web-based medical information [45,46]. One explanation for the low recall scores in this study could be that the RFA information presented was rather complex. However, it was the original RFA information presented on the website of a specialized cancer hospital, and we know that the majority of eHealth information is complex [47]. As message complexity is associated with poor recall of information [46], it is crucial to consider message complexity when designing eHealth materials to improve recall of eHealth information

Third, we did not include the target sample for testing our stimulus materials, that is, cancer patients. Patients might have been more involved and motivated because of their personal experience with cancer and seeking Web-based cancer information. As involvement has often been associated with deeper processing of information [48], including cancer patients might have resulted in higher recall scores than found in the current study. Furthermore, the younger and older adults in our healthy study sample did not differ in their Internet use. Even though older adults are increasingly using the Internet [1], they are still using the Internet considerably less than their younger counterparts [3]. This bias might explain the lack of age differences found in this study. Nevertheless, including healthy individuals still resulted in support for effectively combining audiovisual information with conversational style (vs written information), and these effects might even increase when tested among cancer patients and individuals less experienced with the Internet.

Implications and Future Research Directions

This study adds to the literature by applying modality and narration style strategies to the field of eHealth information. Dealing with eHealth information, especially when it concerns life-threatening diseases such as cancer, often involves feelings of stress and anxiety [49]. In an attempt to reduce such feelings, individuals use different coping strategies [50]. For instance, “monitors” intentionally seek information to reduce stress and “blunters” avoid information to diminish stress [50]. It is recommended to focus on effective communication strategies in the field of eHealth information. Future research could benefit from including cancer patients in such research to examine how coping style interferes with learning material that is emotionally demanding and stressful.

Although the recall scores of both younger and older adults were low, future research should also be aimed at getting more insight into the effectiveness of eHealth information by focusing on what message strategies work best for older audiences in particular. Older adults are often vulnerable to poor Web-based communication, due to their limited eHealth literacy and inexperience with Internet technologies [9]. In the current climate of presenting crucial health information through the Internet, it is important to consider aging populations when designing eHealth information. By better understanding how older adults process Web-based information, more successful eHealth information can be developed based on successful combinations of message characteristics, rather than providing a best guess combination. Previous research has highlighted the importance of recognizing the diversity of message preferences among older adults [51], which should be considered when developing eHealth materials for this older age group. Future research should focus on such (age-related) factors that might impact how eHealth information is remembered to optimize eHealth tools.

Acknowledgments

The Dutch Cancer Society (KWF) and the Amsterdam School of Communication Research/ASCoR, University of Amsterdam, funded this study. KWF and ASCoR management were not involved in the study design, data collection, data analysis, report writing, and decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The formal audiovisual message in which a younger doctor presents information.

[MP4 File (MP4 Video), 17MB - jmir_v17i4e104_app1.mp4]

Multimedia Appendix 2

The formal audiovisual message in which an older doctor presents information.

[MP4 File (MP4 Video), 18MB - jmir_v17i4e104_app2.mp4]

Multimedia Appendix 3

The conversational audiovisual message in which a younger patient presents information.

[MP4 File (MP4 Video), 35MB - jmir_v17i4e104_app3.mp4]

Multimedia Appendix 4

The conversational audiovisual message in which an older patient presents information.

[MP4 File (MP4 Video), 27MB - jmir_v17i4e104_app4.mp4]

References

1. Pernice K, Nielsen J. Web Usability for Senior Citizens: Design Guidelines Based on Usability Studies With People Age 65 and Older. Fremont: Nielsen Norman Group; 2002. URL: http://www.nngroup.com/reports/senior-citizens-on-the-web/ [WebCite Cache ID 6UikzvtVT]
2. File T, Ryan C. Computer and Internet Use in the United States: 2013. 2014. URL: http://www.census.gov/content/dam/Census/library/publications/2014/acs/acs-28.pdf [WebCite Cache ID 6V0dQWPZ6]
3. Statistics Netherlands. ICT gebruik van personen naar persoonskenmerken. [ICT use by persons subject to personal characteristics]. 2014. URL: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71098ned&D1=33,55-59,69-72,93-95,97-100,102-104,106-109,112-113,119,124-133&D2=0,4-6&D3=0,l&HDR=G2,G1&STB=T&VW=T [accessed 2015-03-03] [WebCite Cache ID 6WkyNeWLv]
4. Cresci MK, Jarosz PA, Templin TN. Are health answers online for older adults? Educ Gerontol 2012 Jan;38(1):10-19. [doi: 10.1080/03601277.2010.515890]
5. Xie B. Older adults' health information wants in the internet age: implications for patient-provider relationships. J Health Commun 2009 Sep;14(6):510-524. [doi: 10.1080/10810730903089614] [Medline: 19731124]
6. Iredale R, Mundy L, Hilgart J. An online resource of digital stories about cancer genetics: qualitative study of patient preferences and information needs. J Med Internet Res 2011 Sep 30;13(3):e78. [doi: 10.2196/jmir.1735] [Medline: 22057223]
7. Ziebland S. The importance of being expert: the quest for cancer information on the Internet. Soc Sci Med 2004 Nov;59(9):1783-1793. [doi: 10.1016/j.socscimed.2004.02.019] [Medline: 15312914]
8. Norman CD, Skinner HA. eHealth literacy: Essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
9. Xie B. Older adults, health information, and the internet. Interactions 2008 Jul 01;15(4):44. [doi: 10.1145/1374489.1374499]
10. Kravitz RL, Hays RD, Sherbourne CD, DiMatteo MR, Rogers WH, Ordway L, et al. Recall of recommendations and adherence to advice among patients with chronic medical conditions. Arch Intern Med 1993 Aug 23;153(16):1869-1878. [Medline: 8250648]
11. Ley P. Communicating with patients: improving communication, satisfaction, and compliance. New York: Croom Helm; 1988.

12. Linn AJ, Van Dijk L, Smit EG, Jansen J, Van Weert JCM. May you never forget what is worth remembering: the relation between recall of medical information and medication adherence in patients with inflammatory bowel disease. J Crohns Colitis 2013 Dec 1;7(11):e543-e550 [FREE Full text] [doi: 10.1016/j.crohns.2013.04.001] [Medline: 23660489]
13. Sweller J, Ayres P, Kalyuga S. The modality effect. In: Sweller J, Ayres P, Kalyuga S, editors. Cognitive load theory. New York: Springer; 2011:129-140.
14. Mayer RE, Fennell S, Farmer L, Campbell J. A personalization effect in multimedia learning: Students learn cetter when words are in conversational style rather than formal style. J Educ Psychol 2004;96(2):389-395. [doi: 10.1037/0022-0663.96.2.389]
15. Naik PA, Raman K. Understanding the impact of synergy in multimedia communications. J Market Res 2003 Nov;40(4):375-388. [doi: 10.1509/jmkr.40.4.375.19385]
16. Ginns P. Meta-analysis of the modality effect. Learn Instr 2005 Aug;15(4):313-331. [doi: 10.1016/j.learninstruc.2005.07.001]
17. Ginns P, Martin AJ, Marsh HW. Designing instructional text in a conversational style: a meta-analysis. Educ Psychol Rev 2013 May 7;25(4):445-472. [doi: 10.1007/s10648-013-9228-0]
18. Silliman RA, Troyan SL, Guadagnoli E, Kaplan SH, Greenfield S. The impact of age, marital status, and physician-patient interactions on the care of older women with breast carcinoma. Cancer 1997 Oct 1;80(7):1326-1334. [Medline: 9317187]
19. Jorgensen ML, Young JM, Harrison JD, Solomon MJ. Unmet supportive care needs in colorectal cancer: differences by age. Support Care Cancer 2012 Jun;20(6):1275-1281. [doi: 10.1007/s00520-011-1214-9] [Medline: 21691847]
20. Mayer RE. Multimedia aids to problem-solving transfer. Int J Educ Res 1999 Jan;31(7):611-623. [doi: 10.1016/S0883-0355(99)00027-0]
21. Paivio A. Imagery and verbal processes. New York: Holt, Rinehart, and Winston; 1971.
22. Baddeley A. Working memory. Science 1992 Jan 31;255(5044):556-559. [Medline: 1736359]
23. Mayer RE, Moreno R. Nine ways to reduce cognitive load in multimedia learning. Educ Psychol 2003 Mar;38(1):43-52. [doi: 10.1207/S15326985EP3801_6]
24. Van Gerven PWM, Paas F, Van Merriënboer JJG, Hendriks M, Schmidt HG. The efficiency of multimedia learning into old age. Br J Educ Psychol 2003 Dec;73:489-505. [doi: 10.1348/000709903322591208] [Medline: 14713374]
25. Paas F, Van Gerven PWM, Tabbers HK. The cognitive aging principle in multimedia learning. In: Mayer RE, editor. The Cambridge handbook of multimedia learning. Cambridge, UK: Cambridge University Press; 2005:339-351.
26. Van Gerven PWM, Paas FGWC, Van Merriënboer JJG, Schmidt HG. Cognitive load theory and the acquisition of complex cognitive skills in the elderly: Towards an integrative framework. Educ Gerontol 2000 Sep;26(6):503-521. [doi: 10.1080/03601270050133874]
27. Kreuter MW, Green MC, Cappella JN, Slater MD, Wise ME, Storey D, et al. Narrative communication in cancer prevention and control: a framework to guide research and application. Ann Behav Med 2007 Jun;33(3):221-235. [doi: 10.1080/08836610701357922] [Medline: 17600449]
28. Mayer RE. Multimedia learning. Psychol Learn Motiv 2002;41:85-139. [doi: 10.1016/S0079-7421(02)80005-6]
29. Reinard JC. The empirical study of the persuasive effects of evidence the status after fifty years of research. Hum Commun Res 1988 Sep;15(1):3-59. [doi: 10.1111/j.1468-2958.1988.tb00170.x]
30. McCray AT. Promoting health literacy. J Am Med Inform Assoc 2005;12(2):152-163 [FREE Full text] [doi: 10.1197/jamia.M1687] [Medline: 15561782]
31. Kreuter MW, Holmes K, Alcaraz K, Kalesan B, Rath S, Richert M, et al. Comparing narrative and informational videos to increase mammography in low-income African American women. Patient Educ Couns 2010 Dec;81 Suppl:S6-14 [FREE Full text] [doi: 10.1016/j.pec.2010.09.008] [Medline: 21071167]
32. Grice HP. Logic and conversation. In: Cole P, Morgan JL, editors. Syntax and semantics: Vol. 3, Speech acts. New York: Seminar Press; 1975:51-58.
33. Reichelt M, Kämmerer F, Niegemann HM, Zander S. Talk to me personally: Personalization of language style in computer-based learning. Comput Hum Behav 2014 Jun;35:199-210. [doi: 10.1016/j.chb.2014.03.005]
34. Sparks L, Nussbaum JF. Health literacy and cancer communication with older adults. Patient Educ Couns 2008 Jun;71(3):345-350. [doi: 10.1016/j.pec.2008.02.007] [Medline: 18374536]
35. Kaskutas LA, Graves K. Relationship between cumulative exposure to health messages and awareness and behavior-related drinking during pregnancy. Am J Health Promot 1994;9(2):115-124. [Medline: 10150712]
36. Stanczyk N, Bolman C, Van Adrichem M, Candel M, Muris J, De Vries H. Comparison of text and video computer-tailored interventions for smoking cessation: randomized controlled trial. J Med Internet Res 2014;16(3):e69 [FREE Full text] [doi: 10.2196/jmir.3016] [Medline: 24589938]
37. Jansen J, Van Weert J, Van der Meulen N, Van Dulmen S, Heeren T, Bensing J. Recall in older cancer patients: measuring memory for medical information. Gerontologist 2008 Apr;48(2):149-157. [Medline: 18483427]
38. Wagner JY, Wuensch A, Friess H, Berberat PO. Surgeon-patient communication in oncology. Eur J Cancer Care 2014 Sep;23(5):585-593. [doi: 10.1111/ecc.12177] [Medline: 24393204]
39. Callahan JS, Kiker DS, Cross T. Does method matter? A meta-analysis of the effects of training method on older learner training performance. J Manag 2003 Oct 01;29(5):663-680. [doi: 10.1016/S0149-2063_03_00029-1]

40. Morrow DG, Hier CM, Menard WE, Leirer VO. Icons improve older and younger adults' comprehension of medication information. J Gerontol B Psychol Sci Soc Sci 1998 Jul;53(4):P240-P254. [Medline: 9679516]
41. Meppelink CS, Van Weert JCM, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015 Jan;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]
42. Eastin MS. Credibility assessments of online health information: The effects of source expertise and knowledge of content. J Comput Mediat Commun 2001;6(4). [doi: 10.1111/j.1083-6101.2001.tb00126.x]
43. Rubin DL, Hafer T, Arata K. Reading and listening to oral‐based versus literate‐based discourse. Commun Educ 2000 Apr;49(2):121-133. [doi: 10.1080/03634520009379200]
44. Kessels RPC. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [Medline: 12724430]
45. Bol N, Van Weert JCM, De Haes JCJM, Loos E, De Heer S, Sikkel D, et al. Using cognitive and affective illustrations to enhance older adults' website satisfaction and recall of online cancer-related information. Health Commun 2014;29(7):678-688. [doi: 10.1080/10410236.2013.771560] [Medline: 24160560]
46. Van Weert JCM, Van Noort G, Bol N, Van Dijk L, Tates K, Jansen J. Tailored information for cancer patients on the Internet: effects of visual cues and language complexity on information recall and satisfaction. Patient Educ Couns 2011 Sep;84(3):368-378. [doi: 10.1016/j.pec.2011.04.006] [Medline: 21550757]
47. McInnes N, Haglund BJA. Readability of online health information: implications for health literacy. Inform Health Soc Care 2011 Dec;36(4):173-189. [doi: 10.3109/17538157.2010.542529] [Medline: 21332302]
48. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. Adv Exp Soc Psychol 1986;19:123-205. [doi: 10.1016/S0065-2601(08)60214-2]
49. Farber JM, Weinerman BH, Kuypers JA. Psychosocial distress in oncology outpatients. J Psychosoc Oncol 2008 Oct 18;2(3-4):109-118. [doi: 10.1300/J077v02n03_09]
50. Myers LB, Newman SP, Enomoto K. Coping. In: Kaptein A, Weinman J, editors. Health psychology. United Kingdom: BPS Blackwell; 2004:141-157.
51. Soroka A, Wright P, Belt S, Pham D, Dimov S, Roure D, et al. User choices for modalities of instructional information. Presented at: IEEE International Conference on Industrial Informatics; 2006; Singapore.

Abbreviations

ANCOVA: analysis of covariance
CTML: cognitive theory of multimedia learning
RFA: radio frequency ablation

Edited by G Eysenbach; submitted 22.12.14; peer-reviewed by IR Bruinessen, M Reichelt; comments to author 18.02.15; accepted 03.03.15; published 24.04.15.

Please cite as:
Bol N, van Weert JCM, de Haes HCJM, Loos EF, Smets EMA
The Effect of Modality and Narration Style on Recall of Online Health Information: Results From a Web-Based Experiment
J Med Internet Res 2015;17(4):e104
URL: http://www.jmir.org/2015/4/e104/
doi:10.2196/jmir.4164
PMID:25910416

Original Paper

Access to Care and Use of the Internet to Search for Health Information: Results From the US National Health Interview Survey

Daniel J Amante[1], MPH; Timothy P Hogan[1,2*], PhD; Sherry L Pagoto[3*], PhD; Thomas M English[1*], PhD; Kate L Lapane[4*], PhD

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

[2]Center for Healthcare Organization and Implementation Research (CHOIR), eHealth Quality Enhancement Research Initiative (QUERI), National eHealth QUERI Coordinating Center, Bedford, MA, United States

[3]Division of Preventive and Behavioral Medicine, Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[4]Division of Epidemiology of Chronic Diseases and Vulnerable Populations, Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[*]these authors contributed equally

Corresponding Author:
Daniel J Amante, MPH
Division of Health Informatics and Implementation Science
Department of Quantitative Health Sciences
University of Massachusetts Medical School
368 Plantation Street
Worcester, MA, 01605
United States
Phone: 1 508 856 8480
Fax: 1 508 856 8993
Email: daniel.amante@umassmed.edu

Abstract

Background: The insurance mandate of the Affordable Care Act has increased the number of people with health coverage in the United States. There is speculation that this increase in the number of insured could make accessing health care services more difficult. Those who are unable to access care in a timely manner may use the Internet to search for information needed to answer their health questions.

Objective: The aim was to determine whether difficulty accessing health care services for reasons unrelated to insurance coverage is associated with increased use of the Internet to obtain health information.

Methods: Survey data from 32,139 adults in the 2011 National Health Interview Study (NHIS) were used in this study. The exposure for this analysis was reporting difficulty accessing health care services or delaying getting care for a reason unrelated to insurance status. To define this exposure, we examined 8 questions that asked whether different access problems occurred during the previous 12 months. The outcome for this analysis, health information technology (HIT) use, was captured by examining 2 questions that asked survey respondents if they used an online health chat room or searched the Internet to obtain health information in the previous 12 months. Several multinomial logistic regressions estimating the odds of using HIT for each reported access difficulty were conducted to accomplish the study objective.

Results: Of a survey population of 32,139 adults, more than 15.90% (n=5109) reported experiencing at least one access to care barrier, whereas 3.63% (1168/32,139) reported using online health chat rooms and 43.55% (13,997/32,139) reported searching the Internet for health information. Adults who reported difficulty accessing health care services for reasons unrelated to their health insurance coverage had greater odds of using the Internet to obtain health information. Those who reported delaying getting care because they could not get an appointment soon enough (OR 2.2, 95% CI 1.9-2.5), were told the doctor would not accept them as a new patient or accept their insurance (OR 2.1, 95% CI 1.7-2.5 and OR 2.1, 95% CI 1.7-2.5, respectively), or because the doctor's office was not open when they could go (OR 2.2, 95% CI 1.9-2.7) had more than twice the odds of using the Internet to obtain health information compared to those who did not report such access difficulties.

Conclusions: People experiencing trouble accessing health care services for reasons unrelated to their insurance status are more likely to report using the Internet to obtain health information. Improving the accuracy and reliability of health information resources that are publicly available online could help those who are searching for information due to trouble accessing health care services.

(J Med Internet Res 2015;17(4):e106) doi:10.2196/jmir.4126

KEYWORDS

health services accessibility; health information technology; information-seeking behavior; Patient Protection and Affordable Care Act

Introduction

Background

The passing of the Patient Protection and Affordable Care Act (ACA) of 2010 will significantly increase the number of people insured in the United States by expanding Medicaid, subsidizing insurance for lower-income Americans, and mandating that everyone has coverage or else be forced to pay an additional tax [1]. Although the insurance mandate of the ACA addresses a major barrier to accessing health care services, it may also introduce unintended consequences. *Access to care* is defined as the opportunity to reach and obtain appropriate health care services to satisfy a perceived need [2]. Although access to care will improve for the newly insured, the concept of moral hazard suggests that this increase in insurance will greatly affect the allocation of resources because those who gain coverage will use medical care that they otherwise would not have used if they were required to pay its full cost [3,4]. Taking this predicted increase of health care service utilization into consideration along with the documented shortage of primary care physicians [5] and increase in disease prevalence for the baby boomer generation [6], quickly accessing health care services when needed could become much more difficult in the near future [7]. Despite these substantial implications, little is known about how people who have insurance but are unable to access care in a timely manner are obtaining information to answer their health questions.

The proliferation of the Internet has provided an opportunity for patients to search for and gather medical information, albeit of unknown accuracy and reliability, that was previously unavailable to them. An estimated 74% of adults in the United States go online to access the Internet with up to 80% of them looking for health information online [8]. The Internet can improve health outcomes by increasing the availability of information, providing social support, improving feelings of self-efficacy, and facilitating interaction with the health care system [9,10]. Online health information seeking may also serve as an alternative to more traditional methods of obtaining health information, such as receiving information directly from health care providers, particularly for those who have trouble immediately accessing health care services when needed. Cline and Haynes [11] described 3 ways consumers access online health information: searching directly for health information, participating in support groups, and consulting with health professionals. Patients with access to providers will inevitably use the Internet more to perform functions that involve interaction with their clinical team than those without access to providers. These tasks could include using a health care organization’s Web-based portal to schedule an appointment online, to communicate with a provider over secure messaging, to request prescription refills, or to participate in a telehealth consultation. However, no portal or contact with a health care provider is needed to search the Internet for health information or to use online support groups or health chat rooms to learn more about health topics. Patients who are unable to obtain health care services quickly may turn to such publically available resources on the Internet to address their immediate needs.

Objective

We looked to investigate the relationship between health care access barriers unrelated to insurance coverage and the use of the Internet to retrieve health information. We hypothesized that those who report trouble accessing services will be more likely to search the Internet for health information. To test this hypothesis, we analyzed data from a large, nationally representative sample of noninstitutionalized adults living in the United States. Determining whether patients use the Internet to satisfy a need they are unable to address with traditional health care services could lead to an opportunity to deliver accurate, reliable, and tailored online health resources to certain people. Improving the quality of resources available on the Internet and disseminating them to the appropriate audience could help improve health outcomes, ease frustration due to insufficient access to health care, and reduce expensive use of emergency services.

Methods

Dataset

This study used data from the 2011 National Health Interview Survey (NHIS). The NHIS is a multipurpose health survey conducted by the National Center for Health Statistics (NCHS) and the Centers for Disease Control and Prevention (CDC) annually since 1957 [12]. The NHIS is the principal source of information on the health of the civilian noninstitutionalized household population of the United States. Data are collected through personal household interviews by census interviewers using computer-assisted personal interviewing (CAPI). The NHIS is made up of a core questionnaire that is periodically revised but remains mostly stable year to year. It also contains a supplement questionnaire containing questions on current health topics that vary every year. The version we used (2011) contained supplement questions related to health information technology (HIT) use.

Sample

For each family, 1 child aged 17 years or younger and 1 adult aged 18 years or older were randomly selected to complete the Sample Child and Sample Adult questionnaires. These questionnaires differ slightly and produce 2 different data files. We used the Sample Adult data file for this study. The 2011 NHIS collected data from 33,014 adults aged 18 years or older, with 465 individuals having a proxy answer the questions due to being physically or mentally unable to answer themselves. Individuals who had a proxy answer or who had missing data on any of the variables of interest were excluded from this analysis. The response rate for the Sample Adult survey was 81.60% (33,014/40,458) of eligible adults. The NHIS has oversampled the black population since 1985, the Hispanic population since 1995, and the Asian population since 2006. Additionally, the 2011 NHIS selection process was revised so that black, Hispanic, or Asian individuals who were older than 65 years had an increased chance of being selected as the sample adult for their household. Greater details about the NHIS sampling methodology can be found elsewhere [12].

Measures

Operational Definition of Exposure-Health Service Access Difficulty

The exposure for this analysis was having difficulty accessing health care services or delay getting care for a reason unrelated to insurance status. To define this exposure, we looked at several different questions included in the NHIS. First, we looked at 8 questions asking whether any specific access problems occurred during the past 12 months. These questions were:

1. Did you have any trouble finding a general doctor or provider who would see you?
2. Were you told by a doctor's office or clinic that they would not accept you as a new patient?
3. Were you told by a doctor's office or clinic that they did not accept your health care coverage?
4. Have you delayed getting care because you could not get through on the telephone?
5. Have you delayed getting care because you could not get an appointment soon enough?
6. Have you delayed getting care because once you get there you have to wait too long to see the doctor?
7. Have you delayed getting care because the clinic/doctor's office was not open when you could get there?
8. Have you delayed getting care because you did not have transportation?

All 8 questions had the response options of yes, no, refused, not ascertained, and don't know. Those who refused to answer, an answer was not ascertained, or answered "don't know" on any of the questions (n=338) were excluded from the analysis after determining no significant differences existed between them and those to be included in the analysis.

Operational Definition of Outcome-Health Information Technology Use

The outcome of interest was using the Internet to obtain health information. To define this outcome, we considered 2 questions in the HIT supplement that focused on activities conducted in the preceding 12 months. These questions asked specifically about the use of computers to access the Internet and did not ask about the use of mobile phones and/or tablets. The questions were:

1. During the past 12 months, have you ever used computers to look up health information on the Internet?
2. During the past 12 months, have you ever used computers to use online chat groups to learn about health topics?

From these 2 questions, we categorized 2 mutually exclusive categories of HIT use for the outcome variable. The first category ("HIT use") consisted of those who reported using online chat groups to learn about health topics and/or those who reported using the Internet to search for health information. The second group ("no HIT use") contained all who reported they did not use computers to search for health information on the Internet or online chat rooms to learn about health topics in the previous 12 months.

Additional Covariates

Use of HIT has been found to vary by age [13,14], sex [13,15], race/ethnicity [16], education level [13,14], and marital status [17]. To account for such differences, data were collected on these demographic and socioeconomic covariates. We categorized age into 3 groups (18-35 years, 35-60 years, and ≥60 years), race/ethnicity into 5 groups (non-Hispanic white, Hispanic, Asian, non-Hispanic black, and other/multiple races), education into 4 groups (less than high school, high school diploma/GED, Bachelor's/Associate's degree, and advanced degree), and marital status into 4 groups (married/live with partner, widowed, divorced/separated, and never married). Due to missing household income data, we considered education level as a proxy variable. This allowed for every case in the analysis to have complete data on all variables. People who are sick or have a chronic disease are also more likely to utilize HIT resources [14,18]. To adjust for health status, we looked at 3 variables in the model: self-reported health status (5-point scale from 1=poor to 5=excellent), limitation due to chronic disease (yes/no), and the Charlson Comorbidity Index (CCI) [19]. The CCI values ranged from 0-17 based on the presence of the following medical conditions: myocardial infarction, cerebrovascular disease, chronic pulmonary disease, ulcer disease, cancer, diabetes, renal disease, liver disease, connective tissue disease, and dementia as described in a previous study using NHIS data [20]. A score of zero indicates no comorbidities with higher scores representing a higher predicted risk of mortality attributable to comorbidities [21]. We categorized the CCI scores into 3 groups containing adults with a CCI of 0, 1, and ≥2. Insurance coverage was categorized as having any private insurance coverage, public insurance coverage, or having no insurance coverage.

Analytic Methods

First, weighted percentages within each category of sociodemographic and health status variables separated by HIT use were calculated. Differences within each variable were examined using the Pearson chi-square test for categorical variables. Next, a multinomial logistic regression including all sociodemographic and health status variables was used to calculate adjusted odds ratios (AORs) describing the odds of using HIT compared to the odds of not using HIT within each variable. To accomplish the study objective, multiple multinomial logistic regressions were conducted looking at differences in the odds of using HIT compared to the odds of not using HIT to obtain health information. The first multinomial regression we ran for each exposure variable included only the outcome of interest (HIT use), producing crude ORs for each outcome variable. We then added sociodemographic and health status covariates to each model to account for sex, age, race, education, marital status, health status, and insurance coverage, producing AORs. We chose these covariates because the literature has shown that they significantly differ in respect to HIT use. To account for oversampling, a known nonzero probability of selection for each individual was used in conjunction with adjustments for nonresponse and poststratification to generate sample weights for each individual that were applied to the regressions [12].

Results

Sample Population Characteristics

Of the 33,014 adults surveyed in the 2011 NHIS, 32,139 (97.35%) were included in this analysis. Of those included, 3.63% (1168/32,139) used online health chat rooms to learn about health topics, 43.55% (13,997/32,139) used the Internet to search for health information, and 56.19% (18,059/32,139) did not search the Internet or use online chat rooms to find health information. The majority (92.89%, 1085/1168) of those who used online health chat rooms also reported using the Internet to search for health information in the previous 12 months. Distribution of sociodemographic and health status variables within each category of HIT use are presented in Table 1. All differences were statistically significant (*P*<.001) as determined by the Pearson chi-square test.

Younger participants were more likely to use the Internet to search for health information; 52.84% of participants aged 18-35 years and 51.80% aged 35-60 years reported using either online health chat rooms or the Internet to search for health information compared to only 31.35% of adults aged 60 years or older (data not shown). Women had more than 1.5 times greater odds than men to report using the Internet to search for health information (OR 1.8, 95% CI 1.7-1.9). White participants reported using the Internet to search for health information most frequently (50.01%, 9549/19,095) compared to other races/ethnicities (Hispanic: 30.52%, 1663/5448; Asian: 45.98%, 910/1979; black: 33.44%, 1580/4725; other/multiple races: 42.38%, 378/892). The percentage of adults using HIT to search for health information increased with education level (13.16%, 698/5304 with <high school; 38.94%, 5700/14,637 with high school diploma/GED; 60.47%, 5469/9044 with Bachelor's/Associate's degree; and 70.16%, 2213/3154 with advanced degree). Adults who were widowed used the Internet to search for health information less than those who were married, had never married, or were divorced/separated (18.50%, 549/2968 compared to 48.45%, 7822/16,145; 40.08%, 2201/5492; and 46.56%, 3508/7534; respectively). The odds ratios of using HIT compared to not using HIT for sample characteristics are presented in Table 1.

Reported use of HIT to obtain health information did not change as self-reported health status worsened. Those who reported a limitation due to a chronic condition used the Internet to search for health information less than those who reported no limitation due to a chronic condition (Table 1). Adults with a CCI of zero had lower odds of reporting use of either HIT tool to find health information compared to those with a CCI of 1 (OR 1.3, 95% CI 1.2-1.4) and those with a CCI of 2 or greater (OR 1.4, 95% CI 1.3-1.5). Adults who reported having private insurance coverage had approximately 1.5 times greater odds of using the Internet to search for health information compared to those who were not covered (OR 1.5, 95% CI 1.4-1.6) (Table 1).

Table 1. Sample characteristics by health information technology (HIT) use among adults, NHIS 2011.

Variable	HIT use n=14,080		No HIT use (ref) n=18,059
	%[a]	OR (95% CI)[b]	%[a]
Age (years)			
18-34	34.72	1.0 (ref)	27.48
35-60	48.57	0.7 (0.6-0.7)	40.08
≥60	16.71	0.4 (0.3-0.4)	32.44
Sex			
Men	42.80	1.0 (ref)	53.47
Women	57.20	1.8 (1.7-1.9)	46.53
Race/ethnicity			
Non-Hispanic white	74.81	1.0 (ref)	61.91
Hispanic	9.37	0.6 (0.5-0.6)	17.16
Asian	4.62	0.6 (0.5-0.6)	4.79
Non-Hispanic black	8.83	0.6 (0.5-0.6)	13.67
Other/multiple races	2.47	0.9 (0.8-1.1)	2.47
Education			
<High school	4.98	1.0 (ref)	22.48
High school diploma/GED	41.61	2.8 (2.5-3.2)	51.09
Bachelor’s/Associate’s	38.00	5.7 (5.0-6.4)	21.01
Advanced degree	15.41	9.8 (8.2-11.2)	5.42
Marital status			
Married/live with partner	66.20	1.0 (ref)	56.51
Widowed	2.62	0.5 (0.4-0.5)	8.84
Divorced/separated	10.35	0.8 (0.8-0.9)	12.67
Never married	20.83	0.8 (0.7-0.9)	21.98
Self-reported health status			
Excellent/very good	67.29	1.0 (ref)	54.65
Good	23.57	1.0 (0.9-1.0)	28.59
Fair/poor	9.13	0.8 (0.8-0.9)	16.75
Limited due to chronic condition			
No	88.56	1.0 (ref)	80.32
Yes	11.44	1.0 (0.9-1.1)	19.68
Charlson Comorbidity Index			
0	66.19	1.0 (ref)	60.28
1	18.75	1.3 (1.2-1.4)	19.21
≥2	15.06	1.4 (1.3-1.5)	20.50
Insurance status			
Not covered	13.64	1.0 (ref)	20.17
Private coverage	66.87	1.5 (1.4-1.6)	41.46
Public coverage	19.49	0.9 (0.8-1.0)	38.37

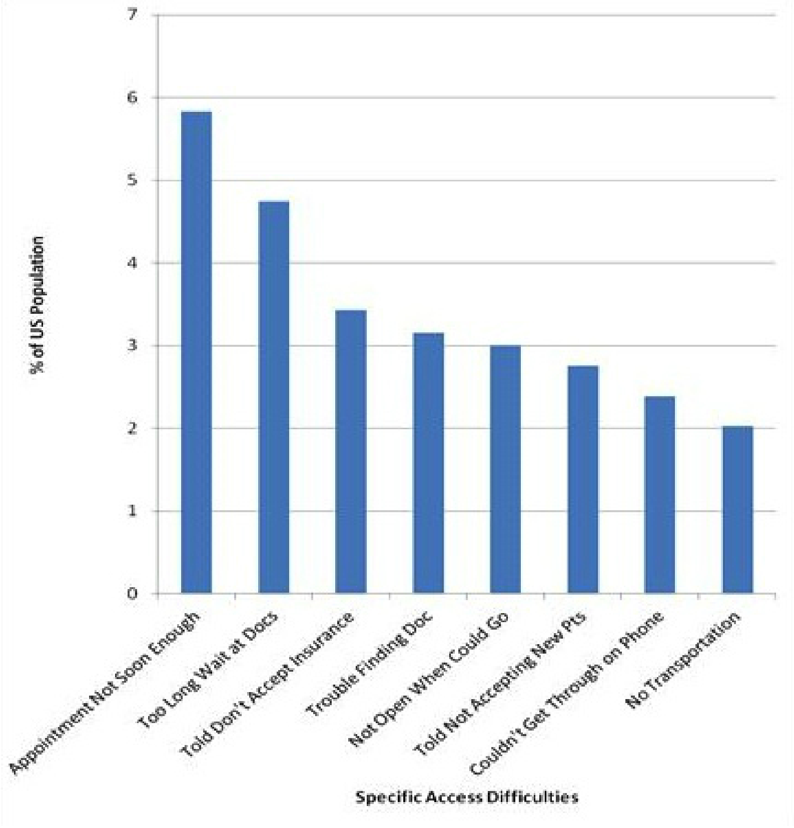
[a] All differences in weighted percentages statistically significant (*P*<.001) as determined by the Pearson chi-square test.

[b] All ORs are adjusted odds of using HIT compared to odds of no HIT use (reference group) accounting for all variables.

Health Care Access Difficulties

Of the 5109 adults who reported experiencing any of the access difficulties in the previous 12 months, 55.47% (2834/5109) reported having experienced only 1 of the 8 specific access difficulties, 22.29% (1139/5109) reported experiencing 2 of the specific access difficulties, 12.92% (660/5109) reported experiencing 3 specific access difficulties, and 9.31% (476/5109) reported experiencing 4 or more of the specific access difficulties in the past 12 months (data not shown). The most commonly reported difficulties accessing health care services in the past 12 months were delaying getting care because they could not get an appointment soon enough (6.09%, 1956/32,139) and delaying getting care because once they got there, they had to wait too long to see the doctor (5.23%, 1681/32,139). The percentage of adults in the US population who reported experiencing each specific type of difficulty are listed in Figure 1.

Figure 1. Percent of US population reporting each access difficulty over the past 12 months.



Relationship Between Health Care Access Difficulty and Health Information Technology Use

The results in Table 2 show the crude and adjusted odds of using HIT to obtain health information for each of the access difficulty questions. Those who reported delays in care because they were told the doctor would not accept them as a new patient or accept their insurance, because they could not get an appointment soon enough, or because the doctor's office was not open when they could go had more than twice (OR 2.1, 95% CI 1.7-2.5; OR 2.1, 95% CI 1.7-2.5; OR 2.2, 95% CI 1.9-2.7; OR 2.2, 95% CI 1.9-2.7; respectively) the odds of reporting use of Internet tools to search for health information after accounting for covariates. Those who reported having trouble finding a doctor who would see them or delayed getting care because they could not get through on the phone or because the wait was too long had more than 1.5 times the odds of reporting use of HIT to search for health information (OR 1.8, 95% CI 1.5-2.1; OR 1.9, 95% CI 1.6-2.4; OR 1.5, 95% CI 1.4-1.8; respectively) after accounting for covariates. Lastly, those who reported delaying care because they did not have transportation were approximately 1.4 times more likely to report use of HIT (OR 1.4, 95% CI 1.1-1.7) after accounting for covariates.

Table 2. Relationship between specific access difficulties over the past 12 months and health information technology (HIT) use.

Specific access difficulty	HIT use n=14,080			No HIT use (ref) n=18,059
	%[a]	OR (95% CI)[b]	AOR (95% CI)[c]	%[a]
Trouble finding a doctor who would see you	3.67	1.38 (1.18-1.60)	1.80 (1.53-2.13)	2.70
Told doctor would not accept you as new patient	3.53	1.73 (1.45-2.06)	2.07 (1.70-2.51)	2.07
Told doctor would not accept your insurance	4.49	1.84 (1.58-2.15)	2.05 (1.71-2.45)	2.49
Delayed getting care because could not get through on phone	3.05	1.71 (1.43-2.04)	1.93 (1.58-2.36)	1.81
Delayed getting care because could not get an appointment soon enough	7.97	2.12 (1.89-2.38)	2.21 (1.94-2.51)	3.93
Delayed getting care because once there, wait was too long to see doctor	5.23	1.22 (1.09-1.38)	1.54 (1.35-1.77)	4.31
Delayed getting care because doctor’s office not open when you could go	4.25	2.30 (1.96-2.69)	2.23 (1.88-2.65)	1.90
Delayed getting care because did not have transportation	1.78	0.78 (0.66-0.93)	1.39 (1.14-1.69)	2.25

[a] Weighted percentage.

[b] Crude ORs are odds of using HIT compared to odds of no HIT use (reference group).

[c] Adjusted ORs accounting for sex, age, race, education, marital status, self-reported health status, presence of chronic disease, CCI, and insurance coverage.

Discussion

This analysis of a large, nationally representative population of adults revealed that more than 15% of US adults reported experiencing difficulty accessing health care services over the previous 12 months for a variety of reasons. The most common difficulties reported were patients delaying getting care because they either could not get an appointment soon enough or because the wait at the doctor’s office was too long. We also found that more than 40% of adults use the Internet to search for health information and approximately 3.7% of adults use online health chat rooms. These results are consistent with previously published studies [8,22].

Who is using the Internet to search for health information and for what purposes needs to be better understood so that online resources can be tailored and improved to maximize its usefulness. Negative consequences of medical Internet use, however, include delivery of inaccurate information, loss of private health information, and potential for harm due to inappropriate or misleading information [23]. Although it has been previously reported that those with higher income [14,22], greater education [13,14,22], of female gender [13-15], younger in age [13,14,24], of nonminority status [16], married [17], and with chronic [14,18] or stigmatized diseases [25] are more likely to use the Internet for health purposes, many of these differences may dissipate after accounting for recent changes in Internet access [15].

As access to the Internet continues to increase across all groups of people, it is important to monitor which patients are turning to the Internet to search for health information, why and how they are using it, and what the resulting effects on health outcomes are. To our knowledge, this study is the first to report that people who experience trouble obtaining needed health services for reasons unrelated to insurance status use the Internet to search for health information more than those who do not have issues accessing services. These findings may have implications for the decreased availability of health care services due to increased insurance coverage resulting from the insurance mandate of the ACA.

Although studies have shown that extending hours of primary care offices reduce the number of emergency room visits [26], a 2007 survey showed that only 28% of adults reported that their regular practice offered hours outside the normal workday or on the weekends, a percentage far less than other industrialized nations [27]. Furthermore, many have expressed concerns that the ACA will further exacerbate the primary care shortage dilemma as citizens who gain public coverage or are mandated to have health coverage begin accessing services at a greater frequency [28], as happened to Massachusetts in 2006 after passing a similar insurance mandated health reform law [29]. As the delays of obtaining primary care appointments continue to increase, the potential for the Internet to help those in need of health information increases as well. For the Internet to be truly helpful, however, people need to know where to find accurate and reliable information, and how to utilize such information to improve their health conditions.

People with a CCI of 1 or more had greater odds of using the Internet for online health chat rooms or to search for health information than those with a CCI of zero. This makes sense because those with comorbidities may rely on the Internet to obtain information about their illnesses. However, people who rated their health as fair/poor had lower odds of searching the Internet for health information compared to those who reported excellent/very good health status. Similarly, those who reported having functional limitation due to a chronic disease also had lower odds of searching the Internet for health information than those who reported no limitation due to a chronic disease. Although previous reports suggest that sicker participants use

HIT tools more often [14,18], including HIT tools provided by their care teams such as online prescription refills, secure messaging, and appointment requests, our analysis was only focused on the publically available HIT tools of searching for health information or online health chat rooms. This limited sample of HIT focus may explain some of the inconsistencies found between our results and the literature, as those who reported poorer health or limitation to chronic conditions could be less likely to search the Internet for health information but more likely to use HIT tools provided by their care team. This study also does not account for Internet access at home. This could explain some of our contradictory findings as previous studies have found that although people with chronic conditions are less likely to have Internet access [30], of those with Internet access, adults with a chronic condition use it more to search for information than those without a chronic condition [18].

We also found that younger people and those with more education have greater odds of reporting use of the Internet to search for health information, consistent with previous studies [13,14,22]. A logical explanation for this is that the health question these adults need answered may not be severe enough, in their opinion, to warrant the full effort or time required to obtain routine, urgent, or emergent clinical attention. Instead of taking time out of their busy day, they may decide to try to resolve the health question on their own. Because this group of adults is more comfortable with the Internet, providing them with easily obtainable, user-friendly, accurate, and reliable online resources could help them make appropriate decisions about how best to maintain or improve their health condition. Additionally, there are providers who offer Web-based video consultations and websites that enable patients to quickly get answers to their health questions from a physician. Although legal and ethical questions regarding compensation, liability, and privacy of personal health information need to be continually addressed [31], this type of innovative delivery of care has potential to improve the accessibility and efficiency of health care services [32]. In addition to increased efficiency and accessibility, a recently published review examined the economic value of clinical telehealth and found that the delivery of home care health services by Web-based video consultations was also cost-effective [33].

This study has several strengths. Because we used data from a nationally representative study, the results produced good estimates of the prevalence of HIT use and the difficulties of accessing health care services in the United States. In addition, our exposure variables (access difficulties) were based on actual examples of problems people have obtaining health care services. This strength is in agreement with the access to care framework suggested by Levesque et al [2], in which access to care is defined as the opportunity to reach and obtain appropriate health care services to satisfy a perceived need. Central to this definition is the process of seeking care. We believe that our exposure questions contain elements of this process and view this agreement as a major strength of our study. The NHIS was also the first nationally representative household survey to collect data on Internet-based HIT use [24]. Furthermore, both the exposure and the outcome variables used the same recall period of 12 months. This increases the meaningfulness of the correlation between the two. Also, due to the wealth of information collected in the survey, we were able to account for several important covariates like race, education, marital status, health status, and insurance coverage. Lastly, since the study population was large and very little data was missing, we were able to include only cases with no data missing from all variables included in the model without needing to exclude a high percentage of adults from the study.

When interpreting the results, it is important to keep the following limitations in mind. The NHIS does not collect data on computer-enabled Internet access at home. Likewise, it also does not monitor Internet access via mobile and tablet devices. Since many low-income and young people may not have access to a computer at home, but may have access to the Internet via mobile phones or tablets, this study may have underestimated HIT use among people who have more difficulty accessing health care services due to financial reasons. We also did not account for differences in HIT use that may occur between patients with not well-understood or stigmatized diseases. Whether the participants were using the Internet to search for health information for themselves or another is an additional limitation. Previous studies have found that almost half of people who go online to search for health information are doing so on behalf of someone else [8]. Lastly, given the large sample population, differences in HIT use should be closely scrutinized to determine if the differences found, while statistically significant, are also meaningful. Since the NHIS is a cross-sectional study, meaningful differences can only be considered as correlational and not causal.

Another important limitation to consider is that the questions we used from the NHIS study only represented a portion of all dimensions that encompass the concept of access to health care. Levesque et al [2] describe 5 different dimensions of access to care: approachability, acceptability, availability and accommodation, affordability, and appropriateness. The majority of the questions we included fall under either the approachability or the availability and accommodation dimensions. No questions in the NHIS contained examples of difficulty accessing services due to dimensions of acceptability, affordability, or appropriateness. Future studies should try to incorporate additional exposure variables that would provide insight into the relationship between difficulty accessing health care services and the use of publically available online health resources.

In conclusion, we found that people who experience difficulty obtaining needed health services use online resources to obtain health information more than those not reporting difficulties obtaining services. With the ACA mandating that all people have health insurance, the patterns of health care utilization will be evolving. If difficulty accessing health care services increase, more patients may turn to resources available to them on the Internet. It needs to be ensured that accurate and reliable resources are available, tailored, and distributed to these people.

Conflicts of Interest

None declared.

References

1. Hofer AN, Abraham JM, Moscovice I. Expansion of coverage under the Patient Protection and Affordable Care Act and primary care utilization. Milbank Q 2011 Mar;89(1):69-89 [FREE Full text] [doi: 10.1111/j.1468-0009.2011.00620.x] [Medline: 21418313]
2. Levesque J, Harris MF, Russell G. Patient-centred access to health care: conceptualising access at the interface of health systems and populations. Int J Equity Health 2013;12:18 [FREE Full text] [doi: 10.1186/1475-9276-12-18] [Medline: 23496984]
3. Buchmueller TC, Grumbach K, Kronick R, Kahn JG. The effect of health insurance on medical care utilization and implications for insurance expansion: a review of the literature. Med Care Res Rev 2005 Feb;62(1):3-30. [doi: 10.1177/1077558704271718] [Medline: 15643027]
4. Manning WG, Marquis MS. Health insurance: the tradeoff between risk pooling and moral hazard. J Health Econ 1996 Oct;15(5):609-639. [Medline: 10164045]
5. Kane GC, Grever MR, Kennedy JI, Kuzma MA, Saltzman AR, Wiernik PH, et al. The anticipated physician shortage: meeting the nation's need for physician services. Am J Med 2009 Dec;122(12):1156-1162. [doi: 10.1016/j.amjmed.2009.07.010] [Medline: 19958898]
6. Martin L, Freedman VA, Schoeni RF, Andreski PM. Health and functioning among baby boomers approaching 60. J Gerontol B Psychol Sci Soc Sci 2009 May;64(3):369-377 [FREE Full text] [doi: 10.1093/geronb/gbn040] [Medline: 19299256]
7. Bodenheimer T, Pham HH. Primary care: current problems and proposed solutions. Health Aff (Millwood) 2010 May;29(5):799-805 [FREE Full text] [doi: 10.1377/hlthaff.2010.0026] [Medline: 20439864]
8. Fox S. Health topics: 80% of internet users look for health information online. Washington, DC: Pew Internet & American Life Project; 2011 Feb 01. URL: http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Health_Topics.pdf [accessed 2015-04-13] [WebCite Cache ID 6XlNxDa04]
9. 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]
10. Ybarra ML, Suman M. Help seeking behavior and the Internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]
11. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
12. Centers for Disease Control and Prevention. NHIS Survey Description. 2012 Jun. URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2011/srvydesc.pdf [accessed 2015-04-13] [WebCite Cache ID 6XlOS6JD2]
13. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
14. Bansil P, Keenan NL, Zlot AI, Gilliland JC. Health-related information on the Web: results from the HealthStyles Survey, 2002-2003. Prev Chronic Dis 2006 Apr;3(2):A36 [FREE Full text] [Medline: 16539777]
15. Brodie M, Flournoy RE, Altman DE, Blendon RJ, Benson JM, Rosenbaum MD. Health information, the Internet, and the digital divide. Health Aff (Millwood) 2000;19(6):255-265 [FREE Full text] [Medline: 11192412]
16. Lorence DP, Park H, Fox S. Racial disparities in health information access: resilience of the Digital Divide. J Med Syst 2006 Aug;30(4):241-249. [Medline: 16978003]
17. Gordon MM, Capell HA, Madhok R. The use of the Internet as a resource for health information among patients attending a rheumatology clinic. Rheumatology (Oxford) 2002 Dec;41(12):1402-1405 [FREE Full text] [Medline: 12468820]
18. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the internet for health information? Health Serv Res 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00510.x] [Medline: 16704514]
19. 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]
20. Meissner HI, Tiro JA, Haggstrom D, Lu-Yao G, Breen N. Does patient health and hysterectomy status influence cervical cancer screening in older women? J Gen Intern Med 2008 Nov;23(11):1822-1828 [FREE Full text] [doi: 10.1007/s11606-008-0775-x] [Medline: 18784967]
21. 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]
22. Atkinson N, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
23. Benigeri M, Pluye P. Shortcomings of health information on the Internet. Health Promot Int 2003 Dec;18(4):381-386 [FREE Full text] [Medline: 14695369]

24. Choi N. Relationship between health service use and health information technology use among older adults: analysis of the US National Health Interview Survey. J Med Internet Res 2011 Apr;13(2):e33 [FREE Full text] [doi: 10.2196/jmir.1753] [Medline: 21752784]
25. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
26. O'Malley AS. After-hours access to primary care practices linked with lower emergency department use and less unmet medical need. Health Aff (Millwood) 2013 Jan;32(1):175-183. [doi: 10.1377/hlthaff.2012.0494] [Medline: 23242631]
27. Schoen C, Osborn R, Squires D, Doty M, Rasmussen P, Pierson R, et al. A survey of primary care doctors in ten countries shows progress in use of health information technology, less in other areas. Health Aff (Millwood) 2012 Dec;31(12):2805-2816. [doi: 10.1377/hlthaff.2012.0884] [Medline: 23154997]
28. Iglehart JK. Health reform, primary care, and graduate medical education. N Engl J Med 2010 Aug 5;363(6):584-590. [doi: 10.1056/NEJMhpr1006115] [Medline: 20660398]
29. Wilson JF. Massachusetts health care reform is a pioneer effort, but complications remain. Ann Intern Med 2008 Mar 18;148(6):489-492. [Medline: 18347363]
30. Fox S, Purcell K. Chronic Disease and the Internet. Washington, DC: Pew Internet & American Life Project; 2010 Mar 24. URL: http://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/ [accessed 2015-04-13] [WebCite Cache ID 6XlOxYL8U]
31. Kluge EW. Ethical and legal challenges for health telematics in a global world: telehealth and the technological imperative. Int J Med Inform 2011 Feb;80(2):e1-e5. [doi: 10.1016/j.ijmedinf.2010.10.002] [Medline: 21067967]
32. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
33. Wade VA, Karnon J, Elshaug AG, Hiller JE. A systematic review of economic analyses of telehealth services using real time video communication. BMC Health Serv Res 2010;10:233 [FREE Full text] [doi: 10.1186/1472-6963-10-233] [Medline: 20696073]

Abbreviations

ACA: Affordable Care Act
CAPI: computer-assisted personal interviewing
CCI: Charlson Comorbidity Index
CDC: Centers for Disease Control and Prevention
HIT: health information technology
NCHS: National Center for Health Statistics
NHIS: National Health Interview Study

Edited by G Eysenbach; submitted 17.12.14; peer-reviewed by YJ Lee, N Choi; comments to author 11.02.15; accepted 18.02.15; published 29.04.15.

Please cite as:
Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL
Access to Care and Use of the Internet to Search for Health Information: Results From the US National Health Interview Survey
J Med Internet Res 2015;17(4):e106
URL: http://www.jmir.org/2015/4/e106/
doi:10.2196/jmir.4126
PMID:25925943

Original Paper

Heart Failure Remote Monitoring: Evidence From the Retrospective Evaluation of a Real-World Remote Monitoring Program

Stephen Agboola[1,2,3*], MD, MPH; Kamal Jethwani[1,2,3*], MD, MPH; Kholoud Khateeb[4], PhD; Stephanie Moore[1,2,3], MD; Joseph Kvedar[1,2,3], MD

[1]Partners Healthcare Center for Connected Health, Connected Health Innovation, Boston, MA, United States
[2]Massachusetts General Hospital, Boston, MA, United States
[3]Harvard Medical School, Boston, MA, United States
[4]Northeastern University, Boston, MA, United States
[*]these authors contributed equally

Corresponding Author:
Stephen Agboola, MD, MPH
Partners Healthcare Center for Connected Health
Connected Health Innovation
25 New Chardon St.
Suite 300
Boston, MA, 02128
United States
Phone: 1 617 643 0291
Fax: 1 617 228 4619
Email: sagboola@mgh.harvard.edu

Abstract

Background: Given the magnitude of increasing heart failure mortality, multidisciplinary approaches, in the form of disease management programs and other integrative models of care, are recommended to optimize treatment outcomes. Remote monitoring, either as structured telephone support or telemonitoring or a combination of both, is fast becoming an integral part of many disease management programs. However, studies reporting on the evaluation of real-world heart failure remote monitoring programs are scarce.

Objective: This study aims to evaluate the effect of a heart failure telemonitoring program, Connected Cardiac Care Program (CCCP), on hospitalization and mortality in a retrospective database review of medical records of patients with heart failure receiving care at the Massachusetts General Hospital.

Methods: Patients enrolled in the CCCP heart failure monitoring program at the Massachusetts General Hospital were matched 1:1 with usual care patients. Control patients received care from similar clinical settings as CCCP patients and were identified from a large clinical data registry. The primary endpoint was all-cause mortality and hospitalizations assessed during the 4-month program duration. Secondary outcomes included hospitalization and mortality rates (obtained by following up on patients over an additional 8 months after program completion for a total duration of 1 year), risk for multiple hospitalizations and length of stay. The Cox proportional hazard model, stratified on the matched pairs, was used to assess primary outcomes.

Results: A total of 348 patients were included in the time-to-event analyses. The baseline rates of hospitalizations prior to program enrollment did not differ significantly by group. Compared with controls, hospitalization rates decreased within the first 30 days of program enrollment: hazard ratio (HR)=0.52, 95% CI 0.31-0.86, *P*=.01). The differential effect on hospitalization rates remained consistent until the end of the 4-month program (HR=0.74, 95% CI 0.54-1.02, *P*=.06). The program was also associated with lower mortality rates at the end of the 4-month program: relative risk (RR)=0.33, 95% 0.11-0.97, *P*=.04). Additional 8-months follow-up following program completion did not show residual beneficial effects of the CCCP program on mortality (HR=0.64, 95% 0.34-1.21, *P*=.17) or hospitalizations (HR=1.12, 95% 0.90-1.41, *P*=.31).

Conclusions: CCCP was associated with significantly lower hospitalization rates up to 90 days and significantly lower mortality rates over 120 days of the program. However, these effects did not persist beyond the 120-day program duration.

(J Med Internet Res 2015;17(4):e101) doi:10.2196/jmir.4417

KEYWORDS

heart failure; telemonitoring; remote monitoring; self-management; hospitalizations; mortality

Introduction

Despite the advances made in the management of heart failure, the burden of disease due to heart failure still remains unacceptably high. Prevalence is projected to increase by 46% from 2012-2030 [1]. Data suggest that hospitalization and mortality rates did not change much from 2000-2010 [2]. Heart failure contributed to about 1 in 9 causes of death in 2009, and it is estimated that about half the 825,000 new cases of heart failure diagnosed annually will die within 5 years of diagnosis [2]. The cost of heart failure is also projected to increase by about 127% from the estimated US $30.7 billion spent in 2012 to US $69.7 billion in 2030 [1].

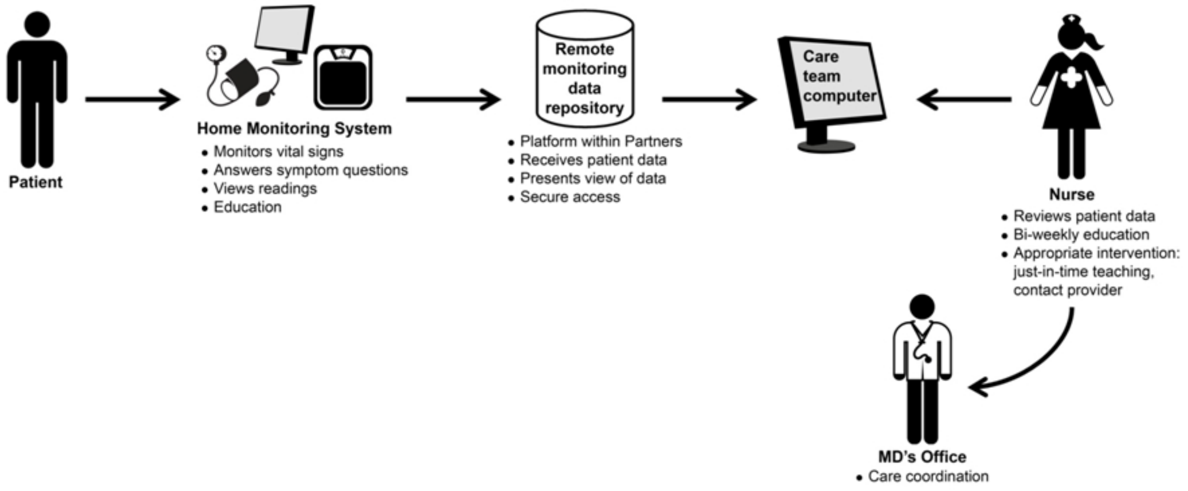
Given the magnitude of this problem, multidisciplinary approaches in the form of disease management programs and other integrative models of care are recommended to optimize treatment outcomes [3]. Remote monitoring, either as structured telephone support or telemonitoring or a combination of both, is fast becoming an integral part of many disease management programs [4-6]. The main strategy is to either provide some sort of education around self-care, or monitor patients for early detection of heart failure decompensation and intervention. Outcomes reported in the current heart failure telemonitoring literature have varied not only based on the type of technology used but also on intensity, complexity of intervention, speed of clinical decision-making, and patient-clinician factors. Telemonitoring approaches vary from simple, noninvasive monitoring of physiologic parameters like heart rate, blood pressure, and weight, to more advanced and invasive approaches like monitoring of intracardiac pressures [7].

An ideal heart failure remote monitoring approach would be one that empower patients for self-care—knowing and able to perform day-to-day care processes, able to early identify symptoms of worsening of disease, and initiating appropriate action. This can be achieved by a care model that incorporates regular out-patient clinic follow-up with ongoing education (provided at clinic visits or via telemonitoring), objective monitoring of clinical condition, and provision of timely feedback. The telemonitoring program at the Partners HealthCare, Center for Connected Health, Connected Cardiac Care Program (CCCP), is one of such programs that model this approach.

CCCP is a 4-month home telemonitoring and education program designed to improve self-management in heart failure patients at risk for hospitalization within the Partners HealthCare network of hospitals. Participants monitor relevant physiologic parameters (blood pressure, heart rate, weight, and blood oxygen saturation) and answer questions on heart failure–related symptoms on a touch-screen computer on a daily basis (Figure 1). The remote monitoring equipment included ViTel Net and devices approved by the Food and Drug Administration: a UA 767PC Turtle 400 monitor, a Life-Source digital weight scale, an A&D blood pressure cuff and meter, and a BCI pulse oximeter device (UC-321PBT). Measurements and responses to symptom questions are transferred securely to a remote monitoring database where the records are reviewed by telemonitoring nurses. Participants also receive structured biweekly telephone-based education sessions over an 8-week period. Patient education covered a variety of topics including diet, physical activity, importance of daily measurements, recognizing symptoms of disease decompensation, and medication adherence. In addition to the structured educational sessions, they received "just-in time" teaching, that is, unscheduled education done to intervene when the remote monitoring nurses observe that measurements fall outside the set baseline range customized for each participant by their physicians or at the onset of new symptoms.

This study aims to evaluate the effect of a heart failure telemonitoring program, Connected Cardiac Care Program (CCCP), on hospitalization and mortality in a retrospective database review of medical records of patients with heart failure receiving care at the Massachusetts General Hospital (MGH). While the heart failure remote monitoring literature is replete with studies in controlled settings, there is a dearth of literature reporting on the evaluation of real-world heart failure remote monitoring programs. In addition, by following up with patients after disenrollment from the CCCP program, this study sheds some light on the downstream impact of completely taking patients (who have built disease self-efficacy skills) off remote monitoring programs.

Figure 1. The Connected Cardiac Care Program.



Methods

Overview

This study is a retrospective analysis to evaluate the effect of CCCP on clinical outcomes in patients enrolled in the program in a 1:1 match cohort study. Matching on potential confounders is a methodology that is commonly adopted to increase efficiency [8,9]. It is generally suitable for situations where the investigators have access to large population data sources [9].

CCCP participants were compared with control patients, within the same health care system, not enrolled in the CCCP. The control patients received the usual standard of care at MGH. A 1:1 individual matching was done to identify controls for each CCCP participant by selecting a control patient that had a hospitalization within 30 days of the corresponding CCCP's patient index hospital admission. Every patient enrolled in the CCCP program must have an index hospitalization. The index admission is the last heart failure–related hospitalization a patient must have prior to enrollment in the CCCP. Other matching parameters are age ±2 years, race, and gender. We used the matching without replacement method. In this method, once a matched patient from the control population is selected, they are no longer eligible for subsequent selection. The best matched control was selected to maximize the precision of the analysis. Gender and race were the first considerations in selecting the best match, followed by the nearest age and index admission date in that order.

Study Population

All subjects included in this study are patients with a diagnosis of heart failure receiving care at MGH. Eligible participants were English-speaking heart failure patients, who had a Partners HealthCare primary care provider or cardiologist that utilized the electronic medical record. They were also required to have a hospital admission to a Partners HealthCare hospital to be eligible to participate in the CCCP. Eligibility requirement for enrollment in the program included that patients must have a diagnosis of heart failure, must have an MGH care provider, and must have been flagged as having a high risk for readmission and referred into the program by their care provider. Patients with end-stage renal disease, on any chemotherapeutic medication, or having had any organ transplant were excluded from participation.

We conducted a retrospective review of the remote monitoring database (RMDR) to identify participants enrolled in the CCCP between Jan. 1, 2008, and Aug. 31, 2012. The RMDR is a secured database housed within the Partners HealthCare firewall where connected health data are processed and stored. From the RMDR, we also collected program information of participants. This information included program start of care and end of care dates. The list of identified program participants was sent to the Partners HealthCare's Research Patient Data Repository (RPDR) to access participants' full clinical data and also to identify eligible match controls. The RPDR is a large clinical data registry that gathers medical records from various hospital systems and stores them in a central location [10]. We identified control patients, not participating in the CCCP but receiving care at MGH, with any heart failure-related International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) and with similar CCCP eligibility requirements. We excluded CCCP patients whose enrollment date in the program was before the year of interest (2008), without documented index hospital admission before CCCP enrollment. We also excluded CCCP patients (and corresponding matched controls) that were enrolled in the program for less than 1 week. Control patients (and corresponding CCCP patient) with incomplete clinical data, those not meeting CCCP entry criteria, and those who died prior to the CCCP enrollment date of corresponding CCCP match were also excluded from analyses.

Source of Data

The primary source of data for this study is the RPDR. The detailed dataset from the RPDR contained demographic and clinical information of patients including age, gender, race, diagnoses, hospital visits, clinical notes, and vital status. All diagnoses were carefully verified in the electronic medical records to limit misclassification of disease status due to data coding errors. We collected and reviewed all-cause hospitalization and mortality data of CCCP patients and their corresponding controls starting from 120 days before enrollment in the CCCP program and up until 1 year after the CCCP enrollment date. The study was approved by the Partners HealthCare Human Research Committee.

Outcome Measures

Outcomes were classified based on follow-up times. The primary outcomes were the effects of the intervention assessed during the program duration of 4 months. The primary effects at 30 days, 60 days, 90 days, and at 4 months were on mortality and hospitalizations.

Secondary outcomes were the cumulative effects of the intervention assessed over a 1-year follow-up period from the date of program enrollment, that is, following up participants for an additional follow-up time of 8-months following program completion. The secondary effects evaluated were on mortality, hospitalizations, risk for multiple hospitalizations, and length of hospital stay.

Statistical Analysis

Baseline data were summarized using descriptive statistics: means and standard deviation for continuous data with normal distribution, medians for skewed data, and percentages for categorical data. We examined group differences using the *t* test or the Wilcoxon-Mann-Whitney test for continuous data and chi-square tests for categorical data. We also assessed baseline hospitalization rates 120 days prior to CCCP enrollment to see if differences existed in hospitalization rates across both groups prior to follow-up. Cumulative survival curves for time-to-event analyses were constructed by the Kaplan-Meier method. The Cox proportional hazard model using time to death and hospitalization as endpoints was used to estimate hazard ratios (HR). Multiple hospitalizations were accounted for in the model. All analyses were stratified on the matched pairs [8], and we also adjusted for age, gender, race, ejection fraction, and New York Heart Association classification (NYHA). A two-sided *P*<.05 was considered as significant. All analyses were performed using data analysis and statistical software, STATA 12 version.

Results

Sample

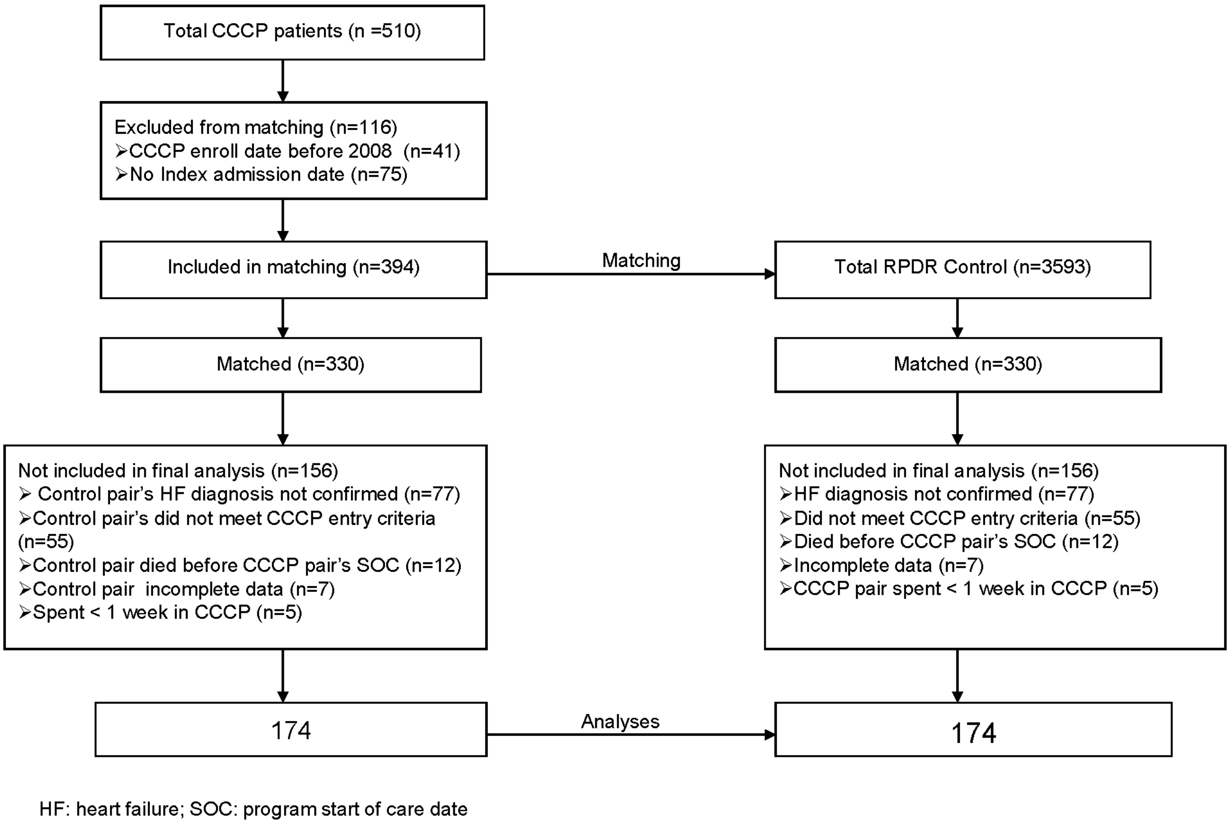
Figure 2 depicts the sample selection process. A total of 510 patients were enrolled in the CCCP from Jan. 1, 2008, to Aug. 31, 2012. Of these, 116 patients were excluded from matching. A total of 348 patients, 174 enrolled in CCCP and 174 match controls, were included in the final analysis.

The majority of the sample population were Caucasian males with an average age of 77 years in both groups, but they differed by marital status. The median duration of follow-up, 365 days, was similar in both groups. The baseline rates of hospitalization, 120 days prior to CCCP enrollment, did not differ significantly in the 2 groups (HR=1.02, 95% CI 0.83-1.24, *P*=.87). All patients were followed up for a maximum duration of 1 year from the period of program enrollment for CCCP patients and their corresponding controls. All baseline characteristics are summarized in Table 1.

Table 1. Baseline characteristics.

Characteristics	CCCP (n=174)	Control (n=174)	*P* value
Male, n (%)	102 (58.62)	102 (58.62)	
Age in years, mean (SD)	76.66 (10.71)	76.76 (10.71)	
Race (white), n (%)	158 (90.80)	158 (90.80)	
Ejection fraction, mean (SD)	49.17 (18.48)	50.66 (16.76)	.43
NYHA class, n (%)			.02
I	5 (2.87)	15 (8.62)	
II	116 (66.67)	125 (71.84)	
III	49 (28.16)	32 (18.39)	
IV	4 (2.30)	2 (1.15)	

Figure 2. The sample selection process.



Primary Outcome

In the model accounting for multiple events, the hospitalization rate in both groups was similar at baseline (HR=1.02, 95% CI 0.83-1.24, *P*=.87). However, after 30 days of enrollment in the program, this rate decreased significantly in the CCCP group compared to the control group (HR=0.52, 95% CI 0.31-0.86, *P*=.01). This differential effect remained consistent throughout the duration of the program: at 60 days (HR=0.66, 95% CI 0.44-0.99, *P*=.05), at 90 days (HR=0.64, 95% CI 0.44-0.99, *P*=.02), and at the end of the 4-month program (HR=0.74, 95% CI 0.54-1.04, *P*=.06). Similar beneficial effect was observed on the mortality rate, which was significantly lower in CCCP patients compared with the control group (HR=0.33, 95% CI 0.11-0.97, *P*=.04) at the end of the program. The number of hospitalization and death events and corresponding rates measured at various points during the follow-up period are reported in Table 2.

Table 2. All-cause hospitalizations and mortality.

Follow-up	CCCP events, n	Control events, n	HR (95% CI)	*P* value
All-cause hospitalizations during and after the program				
120 days before CCCP	220	209	1.02 (0.83-1.24)	.87
30 days	24	49	0.52 (0.31-0.86)	.01
60 days	43	68	0.66 (0.44-0.99)	.05
90 days	56	87	0.64 (0.44-0.99)	.02
120 days (CCCP ends)	75	97	0.74 (0.54-1.02)	.06
1 year	180	151	1.12 (0.90-1.41)	.31
All-cause mortality during and after the program				
120 days (CCCP ends)	5	12	0.33 (0.11-0.97)	.04
1 year	22	31	0.64 (0.34-1.21)	.17

Secondary Outcomes

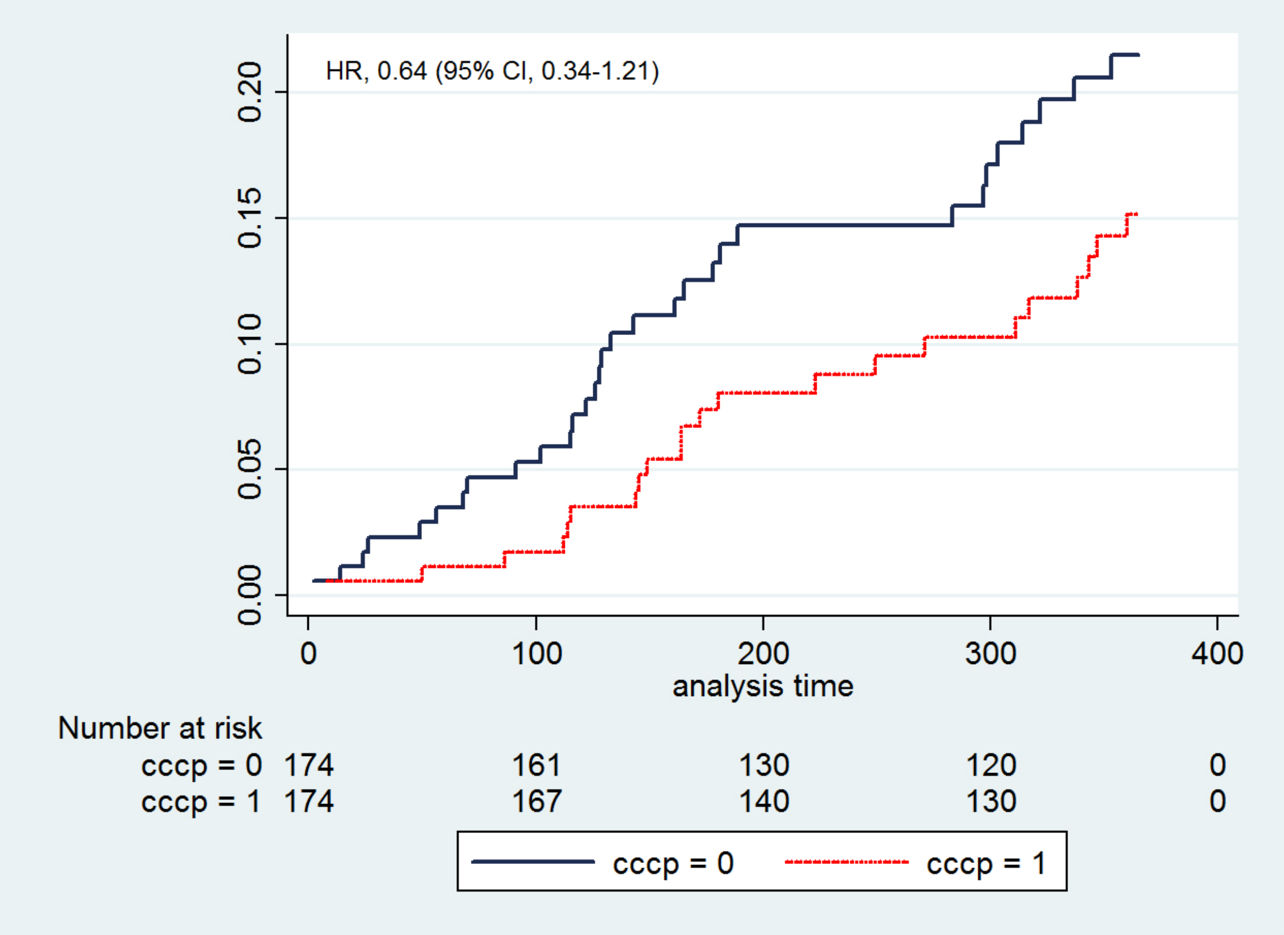
Outcomes did not differ significantly by group over the additional 8 months of follow-up (ie, the course of the 1-year follow-up period from program enrollment). Of the 174 CCCP patients, 47% had at least one all-cause hospitalization over the 1-year follow-up period compared with 46% in controls (relative risk [RR]=1.03, *P*=.83) (Table 3). The risk for more multiple hospitalizations was higher in the CCCP group (25%) in comparison with controls (20%) but was not statistically significant (RR=1.29, *P*=.20). Additionally, the mean length of hospital stay was similar in both groups (Table 3).

Following up patients for an additional 8 months after program completion showed that the differential effect in hospitalization events observed at the end of the program did not persist. The rates increased among the CCCP patients but were not significantly different compared to controls (HR=1.12, 95% CI 0.90-1.41, *P*=.31). Compared with controls, mortality rates over 1-year follow-up were lower in the CCCP group, but this was also not statistically significant (HR=0.64, 95% CI 0.34-1.21, *P*=.17) (Figure 3).

Table 3. Hospitalization events.

Events	CCCP (n=174)	Control (n=174)	RR	*P* value
Any hospitalization, n (%)	82 (47.13)	80 (45.98)	1.03	.83
Multiple hospitalizations, n (%)	44 (25.29)	34 (19.54)	1.29	.20
Length of stay in days, mean (SD)	7 (8.92)	8 (8.83)		.92

Figure 3. All-cause mortality.



Discussion

Principal Findings

This study was designed to evaluate the effect of CCCP, a heart failure remote monitoring program with objective feedback and coaching, compared with matched control patients that received usual care in the similar clinical settings, on clinical outcomes.

Our findings from these retrospective analyses of medical record data suggest that compared to the control group, CCCP was associated with lower hospitalization and mortality rates over the 4-month program duration. Although not statistically significant, the mortality benefit appeared to continue even after following up for an additional 8 months after program completion (ie, 1 year from program enrollment) but hospitalizations increased over this period. We also observed that participants in the program were more likely to have multiple hospitalizations, but there was no difference in length of hospital stay.

Altogether, these findings suggest that the program was associated with reduction in hospitalization and mortality rates during the 4-month program duration and kept patients alive who probably would have died had they not been enrolled in the remote monitoring program. However, the program was associated with increased rates of hospitalizations and mortality after program completion. We speculate that the finding of increased hospitalization and mortality after the 4-month program period could be explained by the progressive nature of the disease, early disproportionate deaths of sicker controls, and also by the fact that participants had become dependent on being monitored remotely with regular access to monitoring devices and the telemonitoring nurses. Tapering of the beneficial effect of monitoring after disenrollment from the program at the end of the 4-month period could also suggest that patients had not developed sufficient self-competency to manage the disease after leaving the program. An alternative explanation for higher hospitalization rates could also be that patients had become sensitized to the early symptoms of disease decompensation due to telemonitoring education and would present earlier than patients who did not have this education.

Although mortality rates increased in the CCCP group after program completion, the overall effect was still beneficial compared to controls over the 1-year follow-up. This finding is similar to results from eight meta-analyses published between 2007 and 2013 evaluating the effect of remote monitoring on mortality [11-17]. These studies reported that compared to usual care, remote monitoring reduced mortality with overall effects ranging from 17% to 51%. The variations in these effects could be explained by the difference in type (structured telephone support vs telemonitoring), speed of feedback (rapid vs non-rapid), invasiveness (invasive vs non-invasive) of remote monitoring, duration of follow-up, study designs, and severity of disease. While a majority of the studies included in these meta-analyses evaluate the effect of remote monitoring only within the monitoring period, this evaluation further monitored participants 8 months beyond the regular 4-month program duration to evaluate the residual effects of the telemonitoring and educational intervention. While the evidence of the association of remote monitoring and reductions in mortality has been consistent over time in meta-analyses, the same effect has not been demonstrated for hospitalizations. The majority of the meta-analyses referenced above reported reductions in hospitalizations except Clarke [14], who did not find a significant reduction in all-cause hospitalization. Likewise, some recent prospective trials have also demonstrated varied effects, ranging from no effect [18-20] to reduction in hospitalizations rates by 37% [21]. Among many other reasons that may account for the incongruity between mortality and hospitalization effects, it is generally easier to assess mortality, which is a hard endpoint that is difficult to miss, unlike hospitalizations, which may be unreported, misclassified, or missed in controls.

Traditionally, remote monitoring is seen as a short-to-medium term adjunct to regular care to empower patients for self-management following hospitalization. Long-term use is not usually feasible due to cost. However, based on our findings, we speculate that increasing the duration of the program to enable patients to develop self-competency may improve outcomes. Although this may not be cost-effective for all participants, risk stratification to identify patients who will benefit from prolonged monitoring may be needed. Alternatively, because patients have built disease self-competency on the program and accompanying monitoring devices, a graduated removal of program components (ie, keeping patients on some sort of less intensive monitoring after discharge from the program) may be helpful. Less intensive monitoring following a remote monitoring program like CCCP is appealing and feasible because of the increasing availability and reducing costs of consumer-oriented monitoring devices that can be easily used by patients [22,23]. Future prospective research to evaluate optimal program monitoring duration, risk stratification to identify patients that may benefit for prolonged monitoring, and the prospects of less intensive, long-term monitoring is needed.

Limitations

Apart from the retrospective nature of this study, it has a number of other limitations. The individual matching done in this study did not include any measure of disease severity. To ensure that patients and their matched controls were comparable at baseline, in terms of severity of disease, we evaluated and found that the rates of hospitalization were similar in both groups 120 days prior to (the CCCP patient and corresponding matched control) enrollment in the CCCP program. We also controlled for measures of disease severity (ejection fraction and NYHA classification) in our analysis. Another limitation is that any hospital admission occurring outside of the Partners' electronic medical records were not captured in these analyses. However, this effect is minimized in the CCCP group because they were monitored daily and had more regular contact with their health care providers. On the other hand, we were more likely to have missed hospitalizations in the control group who may have out-of-system hospitalizations. We also cannot rule out unmeasured confounding from comorbidities and the fact that control patients might have received other treatments including a remote monitoring program other than CCCP. Additionally, given that the program was implemented in the setting of an academic medical center, findings from this evaluation may be generalizable only to such settings.

Conclusions

Results from these analyses suggest that compared with usual care controls, this remote monitoring program is associated with significantly lower hospitalization rates up to 90 days and significantly lower mortality rates over 120 days of the program. However, these effects did not persist over an additional 8 months of follow-up after program completion. There is a need to evaluate the potential impact of risk stratification to determine optimal duration for remote monitoring and also the effect of less intensive, long-term remote monitoring.

Acknowledgments

We would like to acknowledge Regina Nieves, RN, at the Center for Connected Health for her invaluable contribution during the data extraction process.

Conflicts of Interest

None declared.

References

1. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, American Heart Association Advocacy Coordinating Committee, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Cardiovascular Radiology and Intervention, Council on Clinical Cardiology, Council on Epidemiology and Prevention, Stroke Council. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail 2013 May;6(3):606-619 [FREE Full text] [doi: 10.1161/HHF.0b013e318291329a] [Medline: 23616602]
2. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation 2014 Jan 21;129(3):e28-e292 [FREE Full text] [doi: 10.1161/01.cir.0000441139.02102.80] [Medline: 24352519]
3. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, American College of Cardiology Foundation, American Heart Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013 Oct 15;62(16):e147-e239 [FREE Full text] [doi: 10.1016/j.jacc.2013.05.019] [Medline: 23747642]
4. Andrikopoulou E, Abbate K, Whellan DJ. Conceptual model for heart failure disease management. Can J Cardiol 2014 Mar;30(3):304-311. [doi: 10.1016/j.cjca.2013.12.020] [Medline: 24565255]
5. Kulshreshtha A, Kvedar JC, Goyal A, Halpern EF, Watson AJ. Use of remote monitoring to improve outcomes in patients with heart failure: a pilot trial. Int J Telemed Appl 2010;2010:870959 [FREE Full text] [doi: 10.1155/2010/870959] [Medline: 20508741]
6. Agboola S, Havasy R, Myint- U, Kvedar J, Jethwani K. The impact of using mobile-enabled devices on patient engagement in remote monitoring programs. J Diabetes Sci Technol 2013 May;7(3):623-629 [FREE Full text] [Medline: 23759394]
7. Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. Lancet 2011 Aug 20;378(9792):731-739. [doi: 10.1016/S0140-6736(11)61229-4] [Medline: 21856487]
8. Cummings P, McKnight B. Analysis of matched cohort data. Stata Journal 2004;4(3):274-281 [FREE Full text]
9. Sjölander A, Johansson ALV, Lundholm C, Altman D, Almqvist C, Pawitan Y. Analysis of 1 : 1 Matched Cohort Studies and Twin Studies, with Binary Exposures and Binary Outcomes. Statist Sci 2012 Aug;27(3):395-411. [doi: 10.1214/12-STS390]
10. Nalichowski R, Keogh D, Chueh HC, Murphy SN. Calculating the benefits of a Research Patient Data Repository. AMIA Annu Symp Proc 2006:1044 [FREE Full text] [Medline: 17238663]
11. Clark RA, Inglis SC, McAlister FA, Cleland JGF, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. BMJ 2007 May 5;334(7600):942 [FREE Full text] [doi: 10.1136/bmj.39156.536968.55] [Medline: 17426062]
12. Klersy C, De SA, Gabutti G, Regoli F, Auricchio A. A meta-analysis of remote monitoring of heart failure patients. J Am Coll Cardiol 2009 Oct 27;54(18):1683-1694 [FREE Full text] [doi: 10.1016/j.jacc.2009.08.017] [Medline: 19850208]
13. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010;16(2):68-76. [doi: 10.1258/jtt.2009.090406] [Medline: 20008054]
14. Clarke M, Shah A, Sharma U. Systematic review of studies on telemonitoring of patients with congestive heart failure: a meta-analysis. J Telemed Telecare 2011;17(1):7-14. [doi: 10.1258/jtt.2010.100113] [Medline: 21097564]
15. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland JGF. Which components of heart failure programmes are effective? A systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: Abridged Cochrane Review. Eur J Heart Fail 2011 Sep;13(9):1028-1040. [doi: 10.1093/eurjhf/hfr039] [Medline: 21733889]
16. Pandor A, Gomersall T, Stevens JW, Wang J, Al-Mohammad A, Bakhai A, et al. Remote monitoring after recent hospital discharge in patients with heart failure: a systematic review and network meta-analysis. Heart 2013 Dec;99(23):1717-1726. [doi: 10.1136/heartjnl-2013-303811] [Medline: 23680885]
17. Pandor A, Thokala P, Gomersall T, Baalbaki H, Stevens JW, Wang J, et al. Home telemonitoring or structured telephone support programmes after recent discharge in patients with heart failure: systematic review and economic evaluation. Health Technol Assess 2013 Aug;17(32):1-207, v [FREE Full text] [doi: 10.3310/hta17320] [Medline: 23927840]
18. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]

19. Blum K, Gottlieb SS. The effect of a randomized trial of home telemonitoring on medical costs, 30-day readmissions, mortality, and health-related quality of life in a cohort of community-dwelling heart failure patients. J Card Fail 2014 Jul;20(7):513-521. [doi: 10.1016/j.cardfail.2014.04.016] [Medline: 24769270]
20. Vuorinen AL, Leppänen J, Kaijanranta H, Kulju M, Heliö T, van GM, et al. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. J Med Internet Res 2014;16(12):e282 [FREE Full text] [doi: 10.2196/jmir.3651] [Medline: 25498992]
21. Abraham WT, Adamson PB, Bourge RC, Aaron MF, Costanzo MR, Stevenson LW, CHAMPION Trial Study Group. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. Lancet 2011 Feb 19;377(9766):658-666. [doi: 10.1016/S0140-6736(11)60101-3] [Medline: 21315441]
22. Nundy S, Razi RR, Dick JJ, Smith B, Mayo A, O'Connor A, et al. A text messaging intervention to improve heart failure self-management after hospital discharge in a largely African-American population: before-after study. J Med Internet Res 2013;15(3):e53 [FREE Full text] [doi: 10.2196/jmir.2317] [Medline: 23478028]
23. Zan S, Agboola S, Moore SA, Parks KA, Kvedar JC, Jethwani K. Patient engagement with a mobile web-based telemonitoring system for heart failure self-management: a pilot study. JMIR Mhealth Uhealth 2015;3(2):e33 [FREE Full text] [doi: 10.2196/mhealth.3789] [Medline: 25842282]

Abbreviations

CCCP: Connected Cardiac Care Program
ICD-9-CM: International Classification of Disease, Ninth Revision, Clinical Modification
MGH: Massachusetts General Hospital
NYHA: New York Heart Association
RMDR: Remote Monitoring Data Repository
RPDR: Research Patient Data Repository

Edited by G Eysenbach; submitted 10.03.15; peer-reviewed by CP Lau, A Hernández-Madrid; comments to author 03.04.15; revised version received 08.04.15; accepted 09.04.15; published 22.04.15.

Please cite as:
Agboola S, Jethwani K, Khateeb K, Moore S, Kvedar J
Heart Failure Remote Monitoring: Evidence From the Retrospective Evaluation of a Real-World Remote Monitoring Program
J Med Internet Res 2015;17(4):e101
URL: http://www.jmir.org/2015/4/e101/
doi:10.2196/jmir.4417
PMID:25903278

Original Paper

Uncontrolled Web-Based Administration of Surveys on Factual Health-Related Knowledge: A Randomized Study of Untimed Versus Timed Quizzing

Alexander Domnich[1*], MD; Donatella Panatto[1*], PhD; Alessio Signori[1*], PhD; Nicola Luigi Bragazzi[1*], MD, PhD; Maria Luisa Cristina[1*], PhD; Daniela Amicizia[1*], MD; Roberto Gasparini[1*], MD

Department of Health Sciences, University of Genoa, Genoa, Italy

[*]all authors contributed equally

Corresponding Author:
Alexander Domnich, MD
Department of Health Sciences
University of Genoa
Via Pastore, 1
Genoa, 16132
Italy
Phone: 39 010 3538524
Fax: 39 010 3538541
Email: alexander.domnich@gmail.com

Abstract

Background: Health knowledge and literacy are among the main determinants of health. Assessment of these issues via Web-based surveys is growing continuously. Research has suggested that approximately one-fifth of respondents submit cribbed answers, or cheat, on factual knowledge items, which may lead to measurement error. However, little is known about methods of discouraging cheating in Web-based surveys on health knowledge.

Objective: This study aimed at exploring the usefulness of imposing a survey time limit to prevent help-seeking and cheating.

Methods: On the basis of sample size estimation, 94 undergraduate students were randomly assigned in a 1:1 ratio to complete a Web-based survey on nutrition knowledge, with or without a time limit of 15 minutes (30 seconds per item); the topic of nutrition was chosen because of its particular relevance to public health. The questionnaire consisted of two parts. The first was the validated consumer-oriented nutrition knowledge scale (CoNKS) consisting of 20 true/false items; the second was an ad hoc questionnaire (AHQ) containing 10 questions that would be very difficult for people without health care qualifications to answer correctly. It therefore aimed at measuring cribbing and not nutrition knowledge. AHQ items were somewhat encyclopedic and amenable to Web searching, while CoNKS items had more complex wording, so that simple copying/pasting of a question in a search string would not produce an immediate correct answer.

Results: A total of 72 of the 94 subjects started the survey. Dropout rates were similar in both groups (11%, 4/35 and 14%, 5/37 in the untimed and timed groups, respectively). Most participants completed the survey from portable devices, such as mobile phones and tablets. To complete the survey, participants in the untimed group took a median 2.3 minutes longer than those in the timed group; the effect size was small (Cohen’s r=.29). Subjects in the untimed group scored significantly higher on CoNKS (mean difference of 1.2 points, P=.008) and the effect size was medium (Cohen’s d=0.67). By contrast, no significant between-group difference in AHQ scores was documented. Unexpectedly high AHQ scores were recorded in 23% (7/31) and 19% (6/32) untimed and timed respondents, respectively, very probably owing to “e-cheating”.

Conclusions: Cribbing answers to health knowledge items in researcher-uncontrolled conditions is likely to lead to overestimation of people’s knowledge; this should be considered during the design and implementation of Web-based surveys. Setting a time limit alone may not completely prevent cheating, as some cheats may be very fast in Web searching. More complex and contextualized wording of items and checking for the “findability” properties of items before implementing a Web-based health knowledge survey may discourage help-seeking, thus reducing measurement error. Studies with larger sample sizes and diverse populations are needed to confirm our results.

(J Med Internet Res 2015;17(4):e94) doi:10.2196/jmir.3734

KEYWORDS

knowledge questionnaires; online survey; time limit; uncontrolled survey administration; cheating; e-cheating; cribbed answers

Introduction

Measuring people's knowledge of health-related topics in order to assess whether they are sufficiently aware of prevention, medication, and self-care is of particular importance, as health knowledge and health literacy are among the main determinants of health behavior and health status [1,2]. Gaps in people's knowledge, as identified through surveys, can subsequently be targeted by specific information and education interventions. In order to assess people's knowledge of health topics, several tools have been developed in the past few years, including questionnaires, tests, and quizzes on medical/general health [3,4], disease-specific [5-8], and risk factor [9-11] knowledge. Moreover, knowledge items are an essential part of KAP (knowledge-attitude-practices) surveys [12], which are widely used in health care research.

Data collection by means of questionnaires varies in terms of how potential respondents are enrolled, the vehicle used for survey delivery, and the mode of questionnaire administration; all these factors may seriously affect the quality of the data [13]. Web-based surveys have now become important tools in epidemiological and health care data collection and their use seems destined to grow rapidly [14]. Potential advantages and shortcomings of Web-based surveys have been described extensively and reported elsewhere [14-16]. Briefly, the advantages are: the possibility to enroll subjects in distant locations and hard-to-reach and hard-to-involve populations, the automatic nature of the system, which saves researchers time and effort, and the potential cost savings. The disadvantages include uncertainty regarding the quality and validity of the data collected, the representativeness of data, due mainly to the digital divide, and general concerns about the design, implementation, and evaluation of a Web survey [15].

Like other research fields, the assessment of knowledge on health-related topics via Web-based surveys is increasing [3,17-26]. However, since surveys are completed in an environment that is beyond the researcher's control, the Web-based administration of a knowledge questionnaire presents considerable shortcomings. First, a positive response bias is likely to be introduced, since people with more knowledge on a health topic would be more likely to fill in a Web-based survey, which may yield non-representative high scores [27]. For instance, parents with lower nutrition knowledge have been shown to have a higher dropout rate in Web-based assessments of nutrition knowledge, dietary habits, and attitudes of their children [28]. Second, as has been described by Handré et al [29], various social and asocial factors linked to unsupervised survey administration (such as interaction with other people or engagement in activities unrelated to the survey) may have a direct influence on respondents, leading to data contamination. Indeed, these authors found that, among students who were invited to complete a survey in an environment of their choice, 42% were engaged in conversation, one-third had someone present in the room, 21% surfed the Internet, 5.5% received help on questionnaires, and 3.3% changed their answers following someone else's advice. In a recent paper, Jensen and Thomsen [30] examined the influence of nonrandom measurement error on the average level of political knowledge scores obtained in a Web survey. The authors started from the fact that, in comparison with face-to-face or telephone surveys, Web-based surveys tend to record higher knowledge scores, probably as a result of cribbed answers or "e-cheating" (which can be broadly defined as the use of information technology in any type of cheating [31]), when participants are able to interrupt survey compilation and look for correct answers via common search engines. Indeed, they found a high rate of self-reported e-cheating (22.3%) and concluded that this may be an important source of measurement error. Similarly, it has been noted that respondents more frequently give correct answers on cholesterol knowledge items online than in face-to-face interviews [32]. Another potential concern is that participants may respond carelessly owing to lack of interest, in which case their scores are likely to be too low [33,34].

The authors of some earlier studies that used Web-based surveys to assess knowledge of health-related topics through questionnaires administered in unsupervised settings have acknowledged the possibility that respondents might cheat by using additional information sources [3,18,20,25]. In these studies, attempts to prevent cheating were made by ensuring anonymity and encouraging interviewees not to use additional sources. Such attempts, however, may not be efficacious enough.

Although the problem of cheating in Web survey research is recognized, little is known about practical methods of controlling for its effects on data quality, especially in research on health topics. It has been proposed that picture-based items should be used, in order to prevent respondents from using search engines to find the correct answers [35]; such an approach, however, may be difficult to implement in health surveys. Adding specifically designed items to measure self-reported cheating and tabbed browsing should also be considered [30]; however, this strategy risks engendering a social desirability bias, whereby people may underreport engaging in socially undesirable activities [36] such as cheating. Another easy-to-implement method of mitigating the effects of cheating is recording survey completion time and imposing time limits. Malhotra [37] has suggested that survey completion time should always be included as a control variable in statistical models. Indeed, the time taken to complete a Web-based survey can seriously affect data quality [37] in two ways. On the one hand, respondents who rush through a questionnaire may provide less thoughtful [37] or careless answers [33,34]. On the other hand, a long response time may reflect social distraction [29], help-seeking [29,35], or simply greater thought. It has been suggested that setting a time limit may reduce collaboration and help-seeking [38]. The time-based approach seems to be plausible, assuming that cheaters are generally slower than non-cheaters since searching the Web requires some time [30]. However, the effectiveness of time restriction seems to be uncertain, as has been documented in social and political research. Indeed, while Strabac and Aalberg [35] claim that imposing a time limit of half a minute for each answer may solve the problem of cheating, Jensen and Thomsen [30] have demonstrated that e-cheaters are generally quick and that the mean time needed to answer a question is often below 30 seconds.

The present randomized study aimed to compare respondents' performances on a health knowledge survey administered with or without a time limit through an uncontrolled Web-based modality. The subject of the survey was nutrition-related knowledge, a topic of great relevance to public health and one of the most widely studied by means of various data collection modalities. We first hypothesized that respondents who completed the questionnaire within a limited time would score fewer points than those working without a time limit, as they would have less time for help-seeking, social interaction, and e-cheating. Our second hypothesis was that the differences in quiz performance between time-restricted and time-unrestricted groups depend on the type of questionnaire—it would be greater in quizzes more amenable to cheating (primarily as a result of good Web "findability" properties of items). The study did not aim to assess factual nutrition-related knowledge; rather, it constitutes a further step toward finding an optimal modality of conducting Web-based health-related surveys and reducing measurement error.

Methods

Study Setting, Participants, and Procedure

A convenience sample of approximately 150 third-year students at Genoa University (faculties of architecture and education sciences, about 65% females) were recruited for the study in May 2014. The students were told that the aim of the study was to test the feasibility of the Web-based administration of a nutrition-related questionnaire. A short description of the study, the survey instrument, and the modality of survey completion were provided by a researcher from outside the faculty. Students were informed that participation was voluntary, that anonymity was guaranteed, and that the researchers would not know who had filled out the survey. No incentives to participate in the survey were offered. After presentation of the study, the email addresses of those who agreed to participate were collected; all volunteers were able to connect to the Web.

On the day of recruitment, students were randomly allocated into two groups on a 1:1 basis by computer-generated randomization to fill in the questionnaire either without a time limit (TL–) or with a time limit (TL+). Participants were then sent an email containing brief instructions and a direct link to the surveys.

Ethical approval for this anonymous survey was not deemed necessary, since its nature was non-medical and non-interventional; no sensitive data or personal information were collected from volunteers.

Survey Instrument and Outcome Measures

The test consisted of two main parts plus two items on sex and age. There was no "don't know" option, but participants were instructed that they were free to leave out any item if they did not know/were not sure of the correct answer. Only the two items on age and sex were obligatory, as was clearly indicated (by an asterisk). The first part was the validated consumer-oriented nutrition knowledge scale (CoNKS) [11] consisting of 20 true/false items. This tool was chosen because of its brevity, good internal reliability, and criterion and construct validity [11]. CoNKS was translated from English to Italian by means of a back-translation technique by two professional translators. The second part was an ad hoc questionnaire (AHQ) comprising 10 multiple-choice items with 5 response options, one of which was correct (see Multimedia Appendix 1). These 10 items were intentionally designed to be difficult for people without a health care/nutrition background to answer correctly; indeed, they were not intended to measure functional nutrition knowledge, but as a proxy measurement of participants' cheating [39]. At the same time, the AHQ items had good "findability" properties as a result of their "encyclopedic" nature, thus enabling the correct answers to be found easily on the Web. Each of the 10 items was pretested by copying a question into the Google search string; the correct answer could be found directly on the screen among the first 10 search results.

Each correct answer was scored as 1, while incorrect or missing answers were scored as 0. In sum, the resulting total CoNKS score could vary from 0 to 20, while AHQ yielded an overall score out of 10.

In a preliminary study, we had administered the AHQ to 21-23 year old students of engineering (n=15) in an in-class paper-and-pencil researcher-controlled setting. The mean score was 2.3 (SD 1.0) with a range from 1 to 4 points. We therefore assumed that an individual AHQ score of ≥5 in unproctored conditions would have been due to e-cheating.

The survey was implemented by means of professional Web-based survey software QuestionPro. In order to increase respondents' motivation, the layout of both surveys was endowed with a clearly visible progress indicator and all items were scrollable and skippable [40]. Furthermore, this professional software was chosen because of its automatic optimization, such as larger text and easy-to-use buttons, during completion from mobile devices. The layout of both the TL- and TL+ questionnaires was identical, the only difference between the two being the presence of a time limit in the latter case. The time limit of 15 minutes (approximately 30 seconds per question) was imposed in light of previous research [35] and following agreement among research team members that this limit was reasonable. A countdown timer situated next to the progress indicator was clearly visible at the top of the screen. The time taken to complete the survey was automatically recorded by the software. Before clicking on the link, subjects did not know whether or not they would have a time limit to complete the quiz.

The rates of responses and dropouts on both surveys were recorded. The response rate was defined as the number of subjects who started the survey as a proportion of the total number of emails sent (assuming that all emails were read), while the dropout rate was the proportion of subjects who started the survey but did not submit it. Since the QuestionPro software allows multiple entries by the same user to be identified, all submissions were screened for this eventuality. If any subject made multiple entries, all his/her entries were removed from the analysis.

Both links were active for 2 weeks. No reminders were sent, since we were unaware of which students had completed the survey and which had not.

Statistical Analysis

Sample size was computed by means of a two-sided two-sample *t* test. As Dickson-Spillmann et al [11] found a mean difference in CoNKS scores of 1.9 points between subjects with health care or nutrition qualifications and subjects without such qualifications, we supposed that a mean difference of 2 points would have a practical significance (ie, cheating). Therefore, to detect a 2-point difference in mean CoNKS scores between the two groups, with a common SD of 3 points when power is .8 and alpha is .05, at least 36 subjects per group were needed. Considering a non-response rate of 30%, we aimed to enroll 94 subjects.

For descriptive purposes, quantitative variables were expressed as means with SDs or medians with ranges. Descriptive data were expressed as frequencies and percentages with 95% confidence intervals (CIs). To compare categorical data, the χ^2test with Yates correction or Fisher's exact test (in case ≥20% of expected frequencies were <5) were performed. To compare between-group differences in continuous variables, the two-sample Student's *t* test was used when data in each sample showed approximately normal distribution, which was preliminarily assessed by means of both visual inspection and Shapiro-Wilk test; otherwise, the Mann-Whitney *U* test was preferred. The effect size for normally distributed data was measured by means of Cohen's *d* with 95% CIs and U_3 index. Cohen's *d* was interpreted as: small (d=0.2), medium (d=0.5), and large (d=0.8). The effect size for the Mann-Whitney test was expressed as Cohen's $r=z/\sqrt{n}$ and interpreted as follows: small (r=0.1), medium (r=0.3), and large (r=0.5) [41].

To assess whether the impact of time limit group (TL− or TL+) was similar in CoNKS and AHQ, a linear mixed model with interaction between the type of quiz and group was used. Since the two scores were on different scales, the *z*-score transformation was applied first. The linear mixed model enabled us to account for possible correlation between scores obtained by the same subject.

Statistical significance was set at a two-sided *P* value of <.05. All data were analyzed by means of the R stats package, version 3.0.1 [42].

Results

Descriptive Statistics

Of 107 volunteers, 94 (47 per group) were randomized to receive the link to either TL– or TL+. No multiple entries were registered; both surveys were started by 72 individual visitors. Response rates were similar in both groups (75%, 35/47; 95% CI 61-85% and 79%, 37/47; 95% CI 65-89% in TL– and TL+ groups, respectively; χ^2_1=0.06, *P*=.81). The between-group dropout rate did not differ (*P*=.99, Fisher's exact test): 4 participants in the TL– group (11%, 4/35) and 5 in the TL+ group (14%, 5/37) dropped out before submitting responses. Among the 9 dropouts, 7 (78%; 95% CI 44-96%) looked at the survey items without answering any question, while the remaining 2 dropped out after answering some items. Thus, 63 complete surveys (31 and 32 in TL– and TL+ groups, respectively) were analyzed. Overall, both groups were comparable in terms of age, sex, and device used to complete the survey. Notably, 60% (43/72; 95% CI 48-71%) of respondents completed the survey from portable devices (Table 1).

Table 1. Sex, age, and devices used by study participants.

Parameter		TL– (with no time limit)	TL+ (with time limit)	Statistical test
Sex, female[a], n (%; 95% CI)		25/31 (81; 64-92)	26/32 (81; 65-92)	χ^2_1=0.07, *P*=.80
Age, years[a]				t_{61}=1.44, *P*=.16
	mean (SD)	22.2 (1.8)	21.6 (1.7)	
	median (range)	22.0 (20-27)	21.0 (19-26)	
Device used n (%; 95% CI) [b]				Fisher's exact test, *P*=.94
	Desktop/laptop	15/35 (43; 27-60)	14/37 (38; 23-54)	
	Mobile phone	16/35 (46; 30-62)	19/37 (51; 36-67)	
	Tablet	4/35 (11; 4-25)	4/37 (11; 4-24)	

[a] Based on subjects who completed the survey.

[b] Based on subjects who started the survey, since only overall statistics were available.

Outcome Measures

The distribution of survey completion time was substantially right-skewed in both groups (skewness coefficients of 1.6 and 2.0 in TL– and TL+ groups, respectively). TL– group respondents spent more time completing the survey than TL+ group respondents (median 7.8 minutes, range 2.9-30.5 minutes vs median 5.5 minutes, range 3.4-14.7 minutes); the

Mann-Whitney test showed a significant difference (z_U=2.30, *P*=.021). The effect size was, however, judged small (*r*=0.29).

As shown in Table 2, subjects in the TL– group scored significantly higher on CoNKS than those in the TL+ group; Cohen's *d* was medium, being 0.67 (95% CI 0.54-0.80); as shown by the U_3 index, 75% of the TL– group scored above the mean of the TL+ group. By contrast, no statistically significant difference emerged in mean AHQ scores or in the percentage with a score of ≥5. The proportion of non-response items tended to be higher in the TL+ group, especially on CoNKS; the difference, however, did not reach the alpha level of 5% (Table 2).

In the linear mixed model, the group effect (TL– and TL+) on the standardized global score derived from the two questionnaires was statistically significant (*P*=.020), while the interaction between type of questionnaire and group was not (*P*=.19).

Table 2. Participants' performances on the survey instruments.

Questionnaire	Parameter	TL– (with no time limit)	TL+ (with time limit)	Statistical test
CoNKS (consumer-oriented nutrition knowledge scale)				
	Total score, mean (SD)	16.5 (1.9)	15.3 (1.7)	t_{61}=2.73, *P*=.008
	median (range)	17.0 (11-20)	15.5 (12-18)	
	Non-response items, n (%; 95% CI)	1/620 (0.2; 0-0.8)	6/640 (0.9; 0-2)	Fisher's exact test, *P*=.13
AHQ (ad hoc questionnaire)				
	Total score, mean (SD)	3.1 (2.6)	2.6 (1.9)	t_{61}=0.89, *P*=.38
	median (range)	3.0 (0-10)	2.5 (0-7)	
	Score ≥5,n (%; 95% CI)	7/31 (22.6; 11-40)	6/32 (18.8; 8-35)	χ^2_1=0, *P*=.95
	Non-response items, n (%; 95% CI)	23/310 (7.4; 5-11)	29/320 (9.1; 6-13)	χ^2_1=0.37, *P*=.55

Discussion

Key Contributions and Comparison With Prior Work

The present paper contributes to the existing methodological literature on Web-based data collection in epidemiological and health care research in several ways. First of all, the results of our study confirm that the risk of social interactions or e-cheating is likely in Web-based health knowledge surveys since a high proportion of respondents scored unexpectedly high on both quizzes, and this fact must be taken into account during the design and implementation of such studies. Asynchronous communication in time and space and the absence of researcher supervision in Web-based surveys make it extremely difficult to control for cribbed answers in an objective way. We intentionally did not ask participants whether they had used additional information sources or not, since cheating would very probably have been underestimated owing to the social desirability bias; however, we adopted a proxy measure of cheating described in psychological research [39]. We found substantially higher mean CoNKS scores than those recorded by Dickson-Spillmann et al [11] (difference of at least 2.3 points) during CoNKS validation. This superior performance may be partially explained by the relatively high educational level of our participants and the higher proportion of females and the young in our sample; indeed, all these factors have been found to be associated with higher CoNKS scores [11]. However, these sample characteristics can hardly explain the fact that participants in both the TL- and TL+ groups scored even higher than people with health-related or nutrition qualifications in the original study by Dickson-Spillmann et al [11] (16.5/15.3 vs 14.6). Moreover, approximately one-fifth of responders in each group scored unexpectedly high on AHQ, which is very likely due to e-cheating; a similar proportion of e-cheating had been found earlier [30].

Second, our study indicates that using timed surveys in Web-based researcher-uncontrolled assessment of knowledge of health-related topics can mitigate measurement error, as we were able to establish a significant effect of the time limit group on the quiz performance, especially on CoNKS, making our first hypothesis plausible. This finding is also of a certain practical significance, as is shown by the effect size. On the other hand, since the sample size was calculated according to the validated CoNKS survey, the non-significant between-group difference in AHQ scores was likely due to a low statistical power, although participants in the TL+ group tended to score lower. In any case, setting a time limit reduces the median time of survey completion, which, at least virtually, reduces the probability of engaging in survey-unrelated activities and help-seeking. However, imposing a time limit alone is unlikely to prevent help-seeking and e-cheating completely, since some cheaters may be particularly fast in Web searching. Indeed, Jensen and Thomsen [30] have shown that the mean response time on a knowledge item is less than 30 seconds, although the time taken by online respondents in their study varied greatly (relative SDs ranging from 144% to 1243%). High variability in response time is probably due to differences in subjects' Web searching skills. Indeed, subjects in their late teens and twenties have been shown to be very skillful in finding given online contents relatively quickly [43]; this may explain the considerably lower variability in response time among the young adults in our sample.

Third, we suggest that the wording of questions plays an important role in terms of “findability” properties. It may therefore be useful to check whether an item can be easily Googled before undertaking a survey and, if so, to reformulate the question. Indeed, most CoNKS items have “knottier” wording than our AHQ items, and therefore have poorer “findability” properties. For instance, to locate a Web page with the correct information on CoNKS items 4 or 18 (“A healthy meal should consist of half meat, a quarter vegetables, and a quarter side dishes” and “For healthy nutrition, dairy products should be consumed in the same amounts as fruit and vegetables”), a respondent would first have to reflect on a query formulation and then scroll search results, rather than simply copying/pasting a question. To crib correct answers to these types of questions quickly, it is much more advantageous to narrow down the search results by applying an advanced search strategy. However, as demonstrated by Ivanitskaya et al [44], only one-third of students in health sciences are able to use Boolean operators and perform an advanced search. We believe that most e-cheaters in our TL+ group were not able to apply an advanced search strategy and, being under time restriction, scored lower than those in the TL- group. Conversely, as the structure of our AHQ items made them easy to Google (see Methods section), e-cheaters in both groups could find the correct answers relatively quickly, which may explain the absence of a statistically significant between-group difference in AHQ scores. Therefore, quizzes of this type, that is, of a scholastic and encyclopedic character, may be less efficacious in preventing cribbing even if a time limit is set; this finding is in line with that of Jensen and Thomsen [30]. Taken together, these facts may help to understand why we failed to confirm our second hypothesis.

It should be stressed that, if a time limit is set on a knowledge questionnaire, the limit per item or per whole questionnaire should be determined ad hoc, as the probability of e-cheating needs to be balanced against the time needed for cognitive processing. On the one hand, an ample time limit may favor e-cheaters with limited information retrieval capabilities, while on the other hand, as suggested by Jensen and Thomsen [30], it may punish relatively slow non-cheaters; in both cases, the risk of measurement error is high. We suggest that the “30 second condition” [30] per item should not be regarded as a gold standard, as several variables (eg, sociodemographic characteristics of the target population, text readability, survey layout, etc.) may interfere. To identify a more appropriate time limit, pilot researcher-supervised studies would be helpful.

Fourth, more than half of our respondents completed the surveys from portable devices such as mobile phones or tablets. It has been shown that there is little difference between computer and mobile phone administration modes, and survey outcomes assessed by the two modes are generally comparable [45,46], even though a shorter response time from mobile phones than from computers has also been observed [46]. Indeed, the median response time in our survey was substantially below the expected 15 minutes. According to Buskirk and Andrus [46], this finding may be explained by Fitts’s law of human motion, according to which selecting answers by touching the screen of a portable device is quicker than by clicking a mouse (as a result of a shorter distance from the starting point to the target). On the other hand, copying/pasting of a phrase from a touchscreen device is a little more difficult and requires a certain level of manual dexterity, and thus may discourage cheating. We believe that, in our study, these two possible explanations should be seen as complementary rather than mutually exclusive. Moreover, the mobile nature of portable devices allows their owners to complete a survey in various environments; thus, social interaction with other people may also take place, giving rise to another source of measurement error [29,47].

Fifth, the dropout rate in our study (13%, 9/72) was less than half of the 30% rate usually observed in Web-based surveys [48]. This confirms the results of previous research [40,48,49], which indicate that the structure of the survey (eg, a small number of items, absence of grids and graphically complex questions, progress indicator, scrollable and skippable items) can effectively reduce the non-participation bias in population-based studies. By contrast, time restriction is unlikely to interfere with response behavior. Three-quarters of dropouts were “lurkers”, that is, people who view the survey but do not respond to any item [49]. According to Bosnjak and Tuten [49], lurkers are generally motivated to view the survey but not to answer, probably on account of technical difficulties or loss of interest during the survey. It may plausibly be speculated that, in Web-based knowledge surveys, lurkers may contribute to a positive response bias, and therefore to the measurement error. Fortunately, modern software for conducting Web-based surveys enables us to distinguish among various types of non-respondents [49] and thus to better analyze participation patterns.

Limitations

This study probably suffers from participation bias and positive response bias, since many more females than males completed the survey; indeed, not only were more females recruited, but also the participation rate of male students was lower than expected. It has been ascertained that women display higher participation rates in scientific studies [50] and that their participation always tends to be greater in surveys on nutrition knowledge [22-24,26]. However, since the aim of our study was not to evaluate knowledge but to compare the two survey mode conditions, these types of bias have a limited influence on the results. Moreover, application of the simple randomization procedure yielded two comparable groups.

The probability of “lucky guessing”, especially on dichotomous true/false items, should also be acknowledged. While the inclusion of “don’t know” options may discourage guessing [51], Parmenter and Wardle [52] have suggested that such options in nutrition knowledge questionnaires also constitute potential drawbacks, specifically: (1) some respondents who know the correct option but are not confident that it is correct may choose the “don’t know” response, and (2) other subjects, who could work out the correct option if they devoted a little thought to it, might mark “don’t know” as an easier alternative [52]. This is why we decided not to include an explicit “no opinion” response category; however, we tried to discourage guessing by allowing the possibility of leaving questions out.

Conclusions

Cribbed answers to health knowledge items in researcher-unsupervised circumstances are likely. This should be considered during the design and implementation of health knowledge, health literacy, and KAP surveys, and also when comparing results from Web-based questionnaires with those obtained from proctored studies. Subsequent erroneous conclusions and overestimation of health knowledge and health literacy may contribute to poorer health outcomes [53,54]. To date, the only way to prevent cheating is to conduct a face-to-face interview [30] or a pencil-and-paper survey under strict researcher control. However, this may be more time- and resource-consuming than cyberspace surveys [15,32]. The continuous growth of technology use will probably enable novel forms of survey administration. A particularly attractive mode may be the use of synchronous Web-based and voice/video-over-Internet services such as Skype, which enable camera-to-camera video interviews to be conducted [55], thereby preventing cribbing and other forms of cheating. Although research suggests that Skype does not yet have the necessary feasibility characteristics to be used in epidemiological data collection [56], we believe it can already be used in studies targeting the young, since they are the first to have "grown up online" and are quick to adapt to novel technologies [57]. Here, we showed that imposing a time limit may only partially prevent help-seeking and further research is required in order to find an optimal cheating-sensitive vehicle for Internet-based surveys.

Acknowledgments

The study was supported by the Department of Health Sciences – Genoa University, Italy. The authors thank Dr Bernard Patrick for revising the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Ad hoc questionnaire.

[PDF File (Adobe PDF File), 63KB - jmir_v17i4e94_app1.pdf]

References

1. Jadhav A, Andrews D, Fiksdal A, Kumbamu A, McCormick JB, Misitano A, et al. Comparative analysis of online health queries originating from personal computers and smart devices on a consumer health information portal. J Med Internet Res 2014;16(7):e160 [FREE Full text] [doi: 10.2196/jmir.3186] [Medline: 25000537]
2. Sun X, Shi Y, Zeng Q, Wang Y, Du W, Wei N, et al. Determinants of health literacy and health behavior regarding infectious respiratory diseases: a pathway model. BMC Public Health 2013;13:261 [FREE Full text] [doi: 10.1186/1471-2458-13-261] [Medline: 23521806]
3. Tokuda Y, Okubo T, Yanai H, Doba N, Paasche-Orlow MK. Development and validation of a 15-item Japanese health knowledge test. J Epidemiol 2010;20(4):319-328 [FREE Full text] [Medline: 20551582]
4. Nakagami K, Yamauchi T, Noguchi H, Maeda T, Nakagami T. Development and validation of a new instrument for testing functional health literacy in Japanese adults. Nurs Health Sci 2014 Jun;16(2):201-208. [doi: 10.1111/nhs.12087] [Medline: 23991825]
5. López-Silvarrey VA, Pértega DS, Rueda ES, Korta MJ, Iglesias LB, Martínez-Gimeno A. Validation of a questionnaire in Spanish on asthma knowledge in teachers. Arch Bronconeumol 2015 Mar;51(3):115-120 [FREE Full text] [doi: 10.1016/j.arbres.2014.01.015] [Medline: 24656974]
6. Carey MP, Schroder KE. Development and psychometric evaluation of the brief HIV Knowledge Questionnaire. AIDS Educ Prev 2002 Apr;14(2):172-182 [FREE Full text] [Medline: 12000234]
7. Li X, Ning N, Hao Y, Sun H, Gao L, Jiao M, et al. Health literacy in rural areas of China: hypertension knowledge survey. Int J Environ Res Public Health 2013 Mar;10(3):1125-1138 [FREE Full text] [doi: 10.3390/ijerph10031125] [Medline: 23507738]
8. Wright JA, Wallston KA, Elasy TA, Ikizler TA, Cavanaugh KL. Development and results of a kidney disease knowledge survey given to patients with CKD. Am J Kidney Dis 2011 Mar;57(3):387-395 [FREE Full text] [doi: 10.1053/j.ajkd.2010.09.018] [Medline: 21168943]
9. Oncken C, McKee S, Krishnan-Sarin S, O'Malley S, Mazure CM. Knowledge and perceived risk of smoking-related conditions: a survey of cigarette smokers. Prev Med 2005 Jun;40(6):779-784. [doi: 10.1016/j.ypmed.2004.09.024] [Medline: 15850879]
10. Parmenter K, Wardle J. Development of a general nutrition knowledge questionnaire for adults. Eur J Clin Nutr 1999 Apr;53(4):298-308. [Medline: 10334656]
11. Dickson-Spillmann M, Siegrist M, Keller C. Development and validation of a short, consumer-oriented nutrition knowledge questionnaire. Appetite 2011 Jun;56(3):617-620. [doi: 10.1016/j.appet.2011.01.034] [Medline: 21310201]

12. World Health Organization. Advocacy, communication and social mobilization for TB control: a guide to developing knowledge, attitude and practice surveys (WHO/HTM/STB/2008.46). Geneva: World Health Organization; 2008.
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. van Gelder MMHJ, 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]
15. Wright KB. Researching Internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. J Comput Mediated Comm 2005;10(3):00-00. [doi: 10.1111/j.1083-6101.2005.tb00259.x]
16. Evans JR, Mathur A. The value of online surveys. Internet Research 2005 Apr;15(2):195-219. [doi: 10.1108/10662240510590360]
17. Sarmugam R, Worsley A, Flood V. Development and validation of a salt knowledge questionnaire. Public Health Nutr 2014 May;17(5):1061-1068. [doi: 10.1017/S1368980013000517] [Medline: 23507427]
18. Bunting L, Tsibulsky I, Boivin J. Fertility knowledge and beliefs about fertility treatment: findings from the International Fertility Decision-making Study. Hum Reprod 2013 Feb;28(2):385-397 [FREE Full text] [doi: 10.1093/humrep/des402] [Medline: 23184181]
19. Jackson E, Warner J. How much do doctors know about consent and capacity? J R Soc Med 2002 Dec;95(12):601-603 [FREE Full text] [Medline: 12461146]
20. Lauber C, Ajdacic-Gross V, Fritschi N, Stulz N, Rössler W. Mental health literacy in an educational elite -- an online survey among university students. BMC Public Health 2005 May 9;5:44 [FREE Full text] [doi: 10.1186/1471-2458-5-44] [Medline: 15882465]
21. Dissen AR, Policastro P, Quick V, Byrd - Bredbenner C. Interrelationships among nutrition knowledge, attitudes, behaviors and body satisfaction. Health Education 2011 Jun 21;111(4):283-295. [doi: 10.1108/09654281111144256]
22. Kothe EJ, Mullan BA. Perceptions of fruit and vegetable dietary guidelines among Australian young adults. Nutr Diet 2011;68(4):262-266. [doi: 10.1111/j.1747-0080.2011.01550.x]
23. Byrd-Bredbenner C, Wheatley V, Schaffner D, Bruhn C, Blalock L, Maurer J. Development and implementation of a food safety knowledge instrument. J Food Sci Education 2007 Jul;6(3):46-55. [doi: 10.1111/j.1541-4329.2007.00029.x]
24. Ho ASL, Soh NL, Walter G, Touyz S. Comparison of nutrition knowledge among health professionals, patients with eating disorders and the general population. Nutr Diet 2011;68(4):267-272. [doi: 10.1111/j.1747-0080.2011.01549.x]
25. Williams M, Peterson GM, Tenni PC, Bindoff IK. A clinical knowledge measurement tool to assess the ability of community pharmacists to detect drug-related problems. Int J Pharm Pract 2012 Aug;20(4):238-248. [doi: 10.1111/j.2042-7174.2012.00188.x] [Medline: 22775520]
26. Kolodinsky J, Harvey-Berino JR, Berlin L, Johnson RK, Reynolds TW. Knowledge of current dietary guidelines and food choice by college students: better eaters have higher knowledge of dietary guidance. J Am Diet Assoc 2007 Aug;107(8):1409-1413. [doi: 10.1016/j.jada.2007.05.016] [Medline: 17659910]
27. Nagle BJ, Usita PM, Edland SD. United States medical students' knowledge of Alzheimer disease. J Educ Eval Health Prof 2013;10:4 [FREE Full text] [doi: 10.3352/jeehp.2013.10.4] [Medline: 23750313]
28. Vereecken CA, Covents M, Haynie D, Maes L. Feasibility of the Young Children's Nutrition Assessment on the Web. J Am Diet Assoc 2009 Nov;109(11):1896-1902 [FREE Full text] [doi: 10.1016/j.jada.2009.08.013] [Medline: 19857631]
29. Hardre PL, Crowson HM, Xie K. Examining contexts-of-use for web-based and paper-based questionnaires. Educational and Psychological Measurement 2012 Jul 09;72(6):1015-1038. [doi: 10.1177/0013164412451977]
30. Jensen C, Thomsen JPF. Self-reported cheating in web surveys on political knowledge. Qual Quant 2013 Dec 3;48(6):3343-3354. [doi: 10.1007/s11135-013-9960-z]
31. Rogers CF. Faculty perceptions about e-cheating during online testing. J Comp Sci Coll 2006;22(2):206-212.
32. Duffy B, Smith K, Terhanian G, Bremer J. Comparing data from online and face-to-face surveys. Int J Mark Res 2005;47(6):615-639.
33. Ward MK. MS thesis - Graduate Faculty of North Carolina State University. 2014. Using virtual presence and survey instructions to minimize careless responding on Internet-based surveys URL: http://repository.lib.ncsu.edu/ir/handle/1840.16/9362 [accessed 2015-04-07] [WebCite Cache ID 6XbignTWY]
34. Meade AW, Craig SB. Identifying careless responses in survey data. Psychol Methods 2012 Sep;17(3):437-455. [doi: 10.1037/a0028085] [Medline: 22506584]
35. Strabac Z, Aalberg T. Measuring political knowledge in telephone and web surveys: a cross-national comparison. Social Science Computer Review 2010 Jun 30;29(2):175-192. [doi: 10.1177/0894439310371340]
36. Holtgraves T. Social desirability and self-reports: testing models of socially desirable responding. Pers Soc Psychol Bull 2004 Feb;30(2):161-172. [doi: 10.1177/0146167203259930] [Medline: 15030631]
37. Malhotra N. Completion time and response order effects in web surveys. Public Opinion Quarterly 2008 Dec 04;72(5):914-934. [doi: 10.1093/poq/nfn050]
38. Olt MR. Ethics and distance education: strategies for minimizing academic dishonesty in online assessment. Online J Distance Learn Adm 2002;5(3):00-00 [FREE Full text]

39. Lobel TE. Gender differences in adolescents' cheating behavior: an interactional model. Personality and Individual Differences 1993 Jan;14(1):275-277. [doi: 10.1016/0191-8869(93)90204-G]
40. Couper MP, Traugott MW, Lamias MJ. Web survey design and administration. Public Opinion Quarterly 2001 Jun 01;65(2):230-253. [doi: 10.1086/322199]
41. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
42. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. URL: http://www.r-project.org/ [accessed 2015-04-07] [WebCite Cache ID 6Xbiw5Yq8]
43. Hargittai E. Second-level digital divide: differences in people's online skills. First Monday 2002;7(4):00-00. [doi: 10.5210/fm.v7i4.942]
44. Ivanitskaya L, O'Boyle I, Casey AM. Health information literacy and competencies of information age students: results from the interactive online Research Readiness Self-Assessment (RRSA). J Med Internet Res 2006;8(2):e6 [FREE Full text] [doi: 10.2196/jmir.8.2.e6] [Medline: 16867969]
45. Wells T, Bailey JT, Link MW. Comparison of smartphone and online computer survey administration. Social Science Computer Review 2013 Oct 07;32(2):238-255. [doi: 10.1177/0894439313505829]
46. Buskirk TD, Andrus CH. Making mobile browser surveys smarter: results from a randomized experiment comparing online surveys completed via computer or smartphone. Field Methods 2014 Apr 14;26(4):322-342. [doi: 10.1177/1525822X14526146]
47. Mavletova A, Couper MP. Sensitive topics in PC web and mobile web surveys: is there a difference? Surv Res Methods 2013;7(3):191-205 [FREE Full text]
48. Galesic M. Dropouts on the web: effects of interest and burden experienced during an online survey. J Off Stat (Stockh) 2006;22(2):313-328 [FREE Full text]
49. Bosnjak M, Tuten TL. Classifying response behaviors in web-based surveys. J Comput Mediated Comm 2001;6(3):00-00. [doi: 10.1111/j.1083-6101.2001.tb00124.x]
50. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007 Sep;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]
51. Herrmann ES, Heil SH, Sigmon SC, Dunn KE, Washio Y, Higgins ST. Characterizing and improving HIV/AIDS knowledge among cocaine-dependent outpatients using modified materials. Drug Alcohol Depend 2013 Jan 1;127(1-3):220-225 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.07.006] [Medline: 22889696]
52. Parmenter K, Wardle J. Evaluation and design of nutrition knowledge measures. J Nutr Educ 2000 Sep;32(5):269-277. [doi: 10.1016/S0022-3182(00)70575-9]
53. Kelly PA, Haidet P. Physician overestimation of patient literacy: a potential source of health care disparities. Patient Educ Couns 2007 Apr;66(1):119-122. [doi: 10.1016/j.pec.2006.10.007] [Medline: 17140758]
54. Dickens C, Lambert BL, Cromwell T, Piano MR. Nurse overestimation of patients' health literacy. J Health Commun 2013;18 Suppl 1:62-69 [FREE Full text] [doi: 10.1080/10810730.2013.825670] [Medline: 24093346]
55. Janghorban R, Latifnejad RR, Taghipour A. Skype interviewing: the new generation of online synchronous interview in qualitative research. Int J Qual Stud Health Well-being 2014;9:24152 [FREE Full text] [Medline: 24746247]
56. Weinmann T, Thomas S, Brilmayer S, Heinrich S, Radon K. Testing Skype as an interview method in epidemiologic research: response and feasibility. Int J Public Health 2012 Dec;57(6):959-961. [doi: 10.1007/s00038-012-0404-7] [Medline: 22945842]
57. Amicizia D, Domnich A, Gasparini R, Bragazzi NL, Lai PL, Panatto D. An overview of current and potential use of information and communication technologies for immunization promotion among adolescents. Hum Vaccin Immunother 2013 Dec;9(12):2634-2642 [FREE Full text] [doi: 10.4161/hv.26010] [Medline: 23954845]

Abbreviations

AHQ: ad hoc questionnaire
CI: confidence interval
CoNKS: consumer-oriented nutrition knowledge scale
KAP: knowledge-attitude-practices
TL–: without time limit
TL+: with time limit

Edited by G Eysenbach; submitted 25.07.14; peer-reviewed by M Dickson-Spillmann, L Ivanitskaya; comments to author 03.12.14; revised version received 06.12.14; accepted 06.12.14; published 13.04.15.

Please cite as:
Domnich A, Panatto D, Signori A, Bragazzi NL, Cristina ML, Amicizia D, Gasparini R
Uncontrolled Web-Based Administration of Surveys on Factual Health-Related Knowledge: A Randomized Study of Untimed Versus Timed Quizzing
J Med Internet Res 2015;17(4):e94
URL: http://www.jmir.org/2015/4/e94/
doi:10.2196/jmir.3734
PMID:25872617

Viewpoint

Validity of Internet-Based Longitudinal Study Data: The Elephant in the Virtual Room

Carys A Pugh[1], BSc (Hons), MSc; Kim M Summers[1], BSc (Hons), PhD; B Mark C Bronsvoort[1], BVSc, PhD; Ian G Handel[2], BVSc, PhD; Dylan N Clements[2], BVSc, PhD

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom
[2]Royal (Dick) School of Veterinary Studies and The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Carys A Pugh, BSc (Hons), MSc
The Roslin Institute and Royal (Dick) School of Veterinary Studies
University of Edinburgh
Easter Bush
Edinburgh, EH25 9RG
United Kingdom
Phone: 44 1316519167
Fax: 44 1316519105
Email: Carys.Pugh@roslin.ed.ac.uk

Abstract

Background: Internet-based data collection relies on well-designed and validated questionnaires. The theory behind designing and validating questionnaires is well described, but few practical examples of how to approach validation are available in the literature.

Objective: We aimed to validate data collected in an ongoing Internet-based longitudinal health study through direct visits to participants and recall of their health records. We demonstrate that despite extensive pre-planning, social desirability can still affect data in unexpected ways and that anticipation of poor quality data may be confounded by positive validation.

Methods: Dogslife is a large-scale, Web-based longitudinal study of canine health, in which owners of Labrador Retrievers were recruited and questioned at regular intervals about the lifestyle and health of their dogs using an Internet-based questionnaire. The Dogslife questionnaire predominantly consists of closed-answer questions. In our work, two separate validation methodologies were used: (1) direct interviews with 43 participants during visits to their households and (2) comparison of owner-entered health reports with 139 historical health records.

Results: Our results indicate that user-derived measures should not be regarded as a single category; instead, each measurement should be considered separately as each presents its own challenge to participants. We recommend trying to ascertain the extent of recall decay within a study and, if necessary, using this to guide data collection timepoints and analyses. Finally, we recommend that multiple methods of communication facilitate validation studies and aid cohort engagement.

Conclusions: Our study highlighted how the theory underpinning online questionnaire design and validation translates into practical data issues when applied to Internet-based studies. Validation should be regarded as an extension of questionnaire design, and that validation work should commence as soon as sufficient data are available. We believe that validation is a crucial step and hope our suggested guidelines will help facilitate validation of other Internet-based cohort studies.

(J Med Internet Res 2015;17(4):e96) doi:10.2196/jmir.3530

KEYWORDS

epidemiology; validation studies as topic; Internet; questionnaires; longitudinal studies; health; canine

Introduction

The Internet has opened up the field of epidemiology by facilitating public engagement and removing the expense of postal and face-to-face data collection and processing. This particularly relevant for longitudinal studies, which were prohibitively expensive to undertake on a large-scale basis in earlier eras. A corollary of this cost reduction is the need to rely on data generated solely by participants. These data are driven by what investigators hope are well-designed and validated questionnaires.

Re-using questionnaires already known to meet these criteria would minimize effort on the part of investigators. Future meta-analyses would be simplified if free and open access were given to standard questionnaires for different exposures or conditions. Considerable effort has been invested in designing and validating questionnaires that assess exposures, for example those gathering information on diet [1], alcohol intake [2], and smoking [3], and outcomes such as pain [4] and depression [5]. However, for many studies an appropriately tested questionnaire may not exist.

As outlined by Thrusfield [6], questionnaire validity refers to whether the answers truthfully reflect the issues designed to be addressed by the questions. In this paper, we report our experiences of trying to apply the well-understood theory behind data validation (see for example [6,7]) to data collected from members of the UK public. We assess the quality of a newly designed Internet-based questionnaire and use examples to support a series of suggested guidelines for checking the validity of any Web-based longitudinal health study. We report how specific problems such as the timing of checking procedures, the challenges of using user-derived measurements, and the impact of recall decay can affect the quality of data collected. This paper highlights why validation should always be considered as the crucial, final step in questionnaire design.

Methods

Dogslife is the first large scale, Web-based longitudinal study of canine health [8]. In this project, owners of Labrador Retrievers were recruited and questioned at regular intervals about the lifestyle and health of their dogs using an Internet-based questionnaire. In December 2014, the number of participants reached over 5000 owners. The Dogslife questionnaire predominantly consists of closed-answer questions detailed in Multimedia Appendix 1.

Validity of the collected data could be affected at three levels: owner ability to recall incidents of interest, owner understanding of the questionnaire, and owner data entry errors. Quantifying the validity of longitudinal data is challenging as temporal changes (such as a subject growing in height or weight) complicate the simple test-retest scenario. The need to understand and maximize the extent to which Dogslife data reflected the experience of the dogs underlay all future work on the project. Thus, validation of these data had to be undertaken before detailed analysis could begin.

In our work, two separate validation methodologies were used: (1) direct interviews with 43 participants during visits to their households and (2) comparison of owner-entered health reports with 139 historical health records. The details of the sampling methodology and validation process can be found in Multimedia Appendix 2. The results are presented as specific validation problems, with examples of the results obtained. The changes made to the questionnaire in reaction to the findings are described in Multimedia Appendix 1.

Results

Timing of Validation

Overview

The timing of validation depends on the nature of the question being asked. First, early validation assessments allow questionnaire alterations to be instigated as quickly as possible, for example, where closed-answers given may be ambiguous (Example 1). Second, it is important to minimize the delay between a contributor answering the questionnaire and their data being validated. Factors that change with time, such as morphometric measures in growing individuals, are best assessed *before* they have time to change (Example 2). Finally, in the absence of information regarding the rate at which events will be reported, such as for the development of illnesses after the onset of participation, both early and late validation of data may be necessary (Example 3). Early assessment allows obvious reliability issues to be highlighted and immediately corrected, but insufficient events may have occurred to validate all answers. Later assessment will allow time for health records to accrue information for detailed validation (Example 3).

Example 1: Validation of Household Classification

During the visits to participants, it became apparent that participants were unclear as to what age a household member would be regarded as a “child” (Table 1), and owners were inconsistent when deciding whether another pet was “a member of the family” (Table 2). Answer options were revised (extended) with the aim of reducing misclassification (Multimedia Appendix 1). Owners were asked where their dog slept at night and following validation, an explicit option allowing contributors to say that their dog slept with another pet was introduced. After this amendment, the proportion of free-text answers to the sleeping location question dropped from 18.7% to 3.2% (χ^2_1=899, P<2.2x10^{-16}).

Table 1. Comparison of household type in Dogslife record with visit response. Numbers entered refer to the count of households.

Dogslife	Visit response				
	Family, n	More than one adult, n	Retired, n	Single adult, n	Other, n
Family	13	3[a]	0	0	1[b]
More than one adult	0	14	0	0	0
Retired	0	0	6	0	0
Single adult	0	0	2[c]	4	0
Other	0	0	0	0	0

[a] Using “family” as a descriptor in the categories allowed owners to make their own judgment regarding what comprised a family. The intention had been to capture households including children under 16 years of age. Two of these visited households included children older than 16 years of age and the third was a married couple without children who regarded themselves as a “family”, rather than being “more than one adult”.

[b] This household comprised a couple who had children staying with them every weekend.

[c] Both households contained single, retired adults. Our categories were not mutually exclusive.

Table 2. Comparison of sleeping location in Dogslife record with visit response. Numbers entered refer to the count of households.

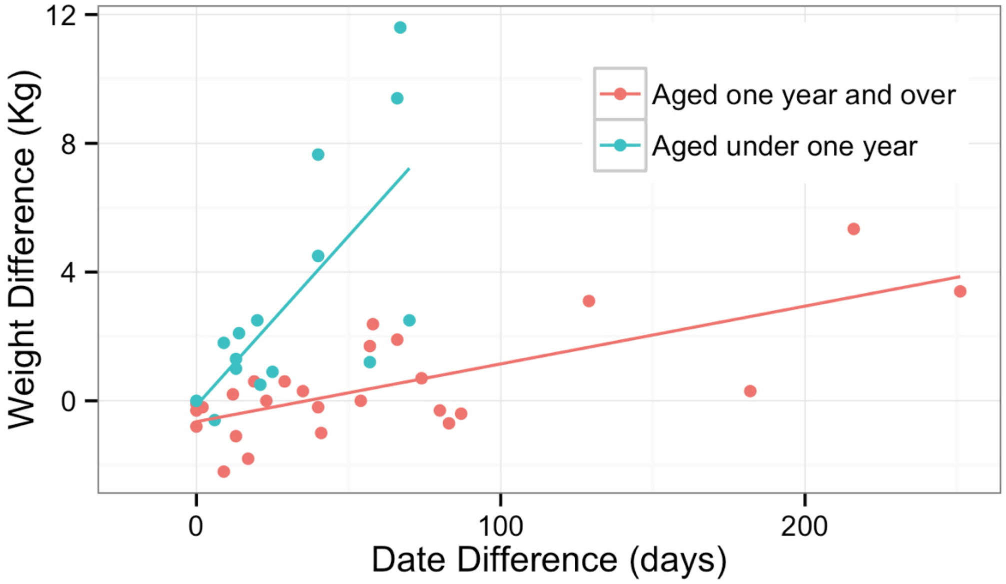
Dogslife	Visit response			
	Alone, n	Shared (family), n	Other (shared with dog)[a], n	Other, n
Alone	24	1	2	0
Shared (family)	0	7	0	0
Other (shared with dog)[a]	0	0	8	0
Other	0	0	1	0

[a] Other (shared with dog) was not an option in the original questionnaire. During the validation, it became apparent that dogs that slept with other dogs were categorized inconsistently as (1) Alone, (2) Shared (family), and (3) Other: (with other dog). Introducing sharing with another pet as an explicit option prevents this inconsistency.

Example 2: Validation of Weight

Weight was compared between the online record and the validation visits. The delay between participant contribution and subsequent validation, and the difference in weight measurements taken by the contributor and the investigator who took validating measurements are shown in Figure 1. The time that had elapsed between the owner completing the questionnaire online and the validation visit (where the investigator took the measurements) was typically too long.

Figure 1. Weight misclassification compared to the time delay between online entry and validation visit (blue data points refer to dogs under 1 year of age and the red to those over 1 year of age, both with linear regression lines).



Example 3: Validation of Health Records

Validation of owner-recorded health data by comparison to veterinary records was a two-stage process (Multimedia Appendix 2). Checking that the questionnaire addressed commonly occurring illnesses was possible at an early stage when little data had accumulated. However, a more quantitative assessment of recall decay and efficiency of reporting illnesses necessitated a delay to allow for sufficient events to have occurred.

Data Not Validated

Overview

The lack of an external source for validation for certain data (eg, illnesses that do not present to a health care professional, or the use of drugs that can be obtained from multiple sources; see Example 4) highlights the potential value of such data. The information is unavailable elsewhere, but its value must be tempered by awareness that there is no robust way of assessing its accuracy.

Example 4: Validation of Over-the-Counter Medication

We tried to ascertain the accuracy of data collected about the timing of administration and products used for endo- and ectoparasitic control (worm and flea treatments). However, neither could be validated from veterinary records because these products can be purchased from multiple sources away from a veterinary surgery, and thus the veterinary record is not an accurate measure of their usage. Furthermore, during the participant visits, so few owners had these products at-hand that comparisons of brand with those entered in the record were impossible.

Pitfalls and Unexpected Benefits of Validation

Overview

Although investigators may anticipate that certain data are more or less likely to be validated during the validation procedure, on the basis of their preconceptions and past experiences, such premonition may be misleading. The results of validation procedures revealed that the dietary information collected was representative for simple classifications, such as the type of food consumed by participating dogs. However, as was expected, it did not capture the full complexity of their diets (Example 5). Conversely, the expectation that other types of information might be challenging to collect accurately (such as dog weight) was confounded by the results (Example 6). Simply because one perceives a question to be difficult to answer should not preclude it from both the questionnaire and validation procedure.

Example 5: Validation of Diet

The Dogslife records and visit responses for simple classification of diet, such as whether it contained dried or a mixture of dried and tinned food, had good agreement (35/39 correct; see Table 3). Unfortunately agreement was poorer with more complicated diets, such as dried food with varying home-prepared additions. The reported food weight measurements were comparatively reliable for wet food but were of varying quality for the other food types where the weight could not be read from the side of a packet, indicating that owners were not weighing the food themselves. In some cases, the owner had documented the weight of a single meal, rather than the whole daily quantity of food (specifically requested in the questionnaire). Validation was possible but the process highlighted that detailed diet was not well captured.

Table 3. Comparison of food types in Dogslife record with visit response. Numbers entered refer to the count of households.

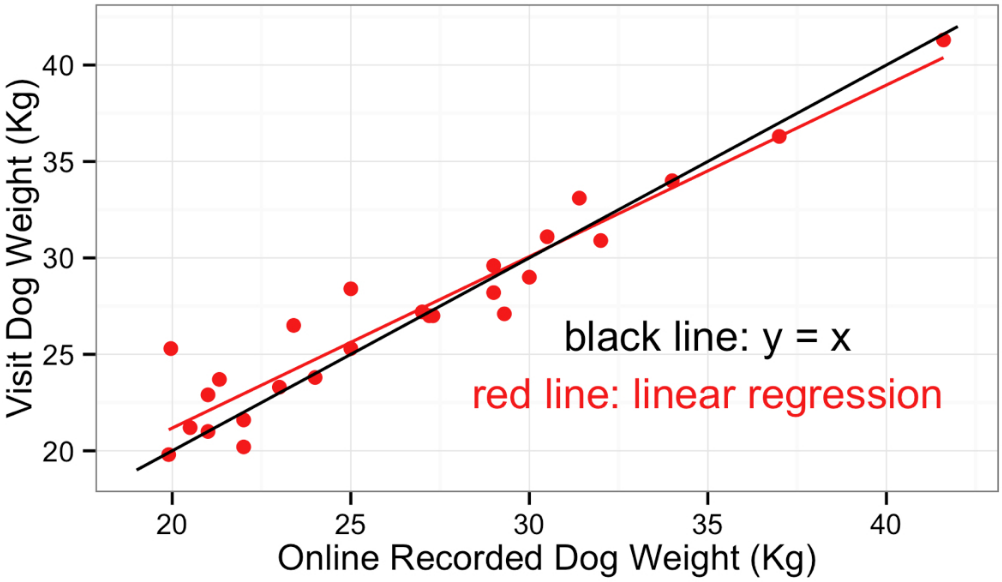
Dogslife	Visit response				
	Dried, n	Tinned, n	Dried & tinned, n	Home prepared, n	Other, n
Dried	27	0	0	1[a]	3[a]
Tinned	0	0	0	0	0
Dried & tinned	0	0	7	0	0
Home prepared	0	0	0	1	0
Other	0	0	0	0	4

[a] Four owners described their dog's diets as "Dried" online but elaborated in person to describe a diet of dried proprietary dog food, with the addition of meat, vegetables, rice, gravy, and fruit.

Example 6: Validation of Weight Measures

A comparison of online dog weight measurements for dogs over 1 year and the weights measured by the investigator during visits for the dogs over 1 year is shown in Figure 2. Lin's concordance correlation coefficient showed moderate to substantial agreement between the two measures at .95 (95% CI 0.89-0.98) [9,10]. During the design stage of the questionnaire, it was decided that dog weight would not be a mandatory question because of perceived difficulty of measurement. Not only did validation indicate that weights given were representative, but 84% (95% CI 83.1-84.2) of data entries included a dog weight and many owners mentioned making special trips to the vet in order to have their dog weighed for the project.

Figure 2. Comparison of online reported dog weight and visit measured dog weight for dogs over 1 year of age.



Social Desirability

Pitfalls and Unexpected Benefits of Validation

Subtle perception and societal pressures may have promoted misclassification for some questions either in the online questionnaire or during the validation visits. For example, the smoking status of household members (a seemingly simple question) was underestimated in the online questionnaire (Example 7), and this may have been due to social perception that smoking is undesirable.

Example 7: Validation of Smoking Status

Smoking status is addressed during the registration phase of the Dogslife study and was validated during owner visits (Table 4).

The overall kappa score for agreement of smoking status was a moderate at .48 (SE=0.19), and there was a significant degree of misclassification among the low number of smokers. During visits, evidence of cigarette smoking was observed in the houses of people who had originally described themselves as non-smokers. With prompting, they said that they only smoked outside. This type of misclassification may have been due to the phrasing of the question: “Does anybody in the household smoke?”, which was perhaps interpreted as “Does anybody smoke within the household/house?”, thus classifying those who only smoke outside as non-smokers.

Table 4. Comparison of smoking in Dogslife record with visit response. Numbers entered refer to the count of households.

Dogslife	Visit response	
	Smoking, n	Non-smoking, n
Smoking	3	1
Non-smoking	4	33

Recall Decay

Overview

We observed recall decay to differing degrees across the dataset through our validation procedure (Example 8), both with vaccinations (Figure 3) and illness reporting (Figure 4).

Figure 3. Vaccination recall decay plot (293 vaccinations, 127 dogs). Those not reported to Dogslife were vaccinations found in the veterinary record but omitted from the Dogslife questionnaire answers.

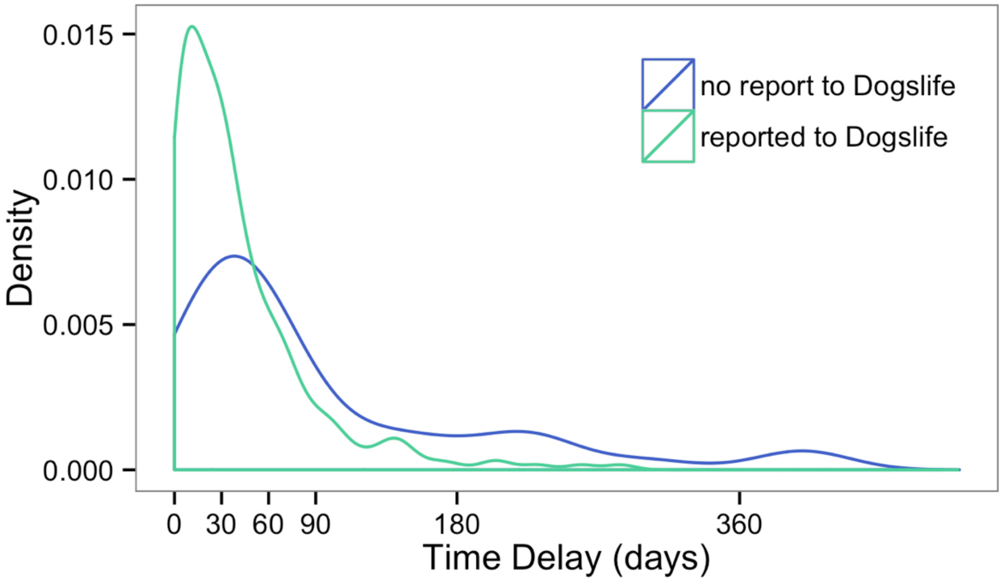
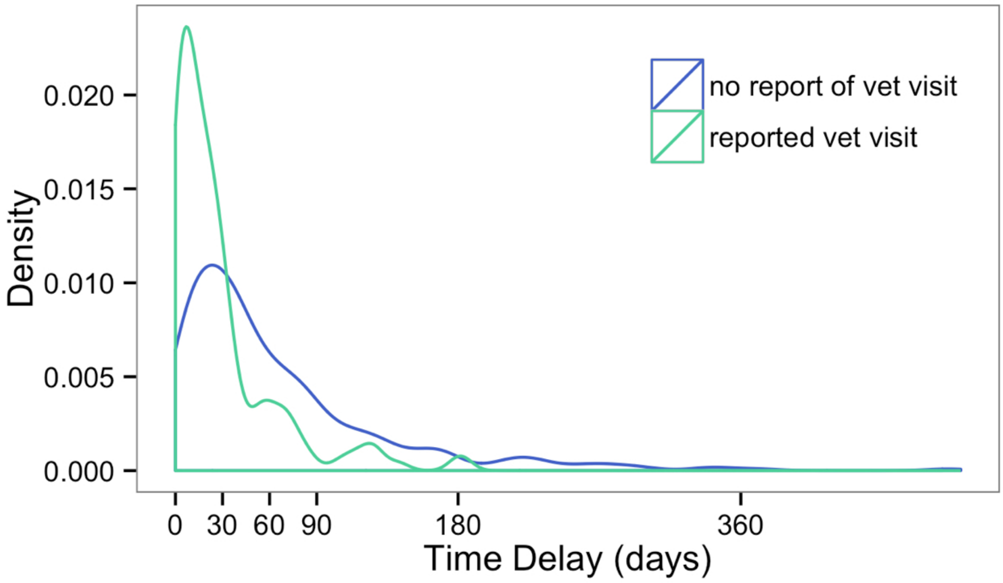


Figure 4. Illness recall decay plot (493 veterinary visits, 101 dogs). Those not reported to Dogslife were illnesses found in the veterinary record but omitted from the Dogslife questionnaire answers (22%).



Example 8: Recall Decay of Vaccinations and Illness Reporting

A clear inverse association of the delay between vaccination and subsequent Dogslife data entry, and likelihood of reporting a vaccination was noted. Questionnaires that incorrectly omitted vaccinations were associated with a greater time delay (Figure 3). Participants may answer the questionnaire at any time, but recent analysis of the cohort return intervals shows a largely bimodal distribution with peaks at 37 days for owners of dogs under 1 year and 90 days, which is the requested time for 3-monthly data entries for dogs aged over 1 year. As such, the delay between an incident and the subsequent data entry (triggered by the reminder email) introduces the potential for recall decay.

Reporting of illnesses also appeared to be affected by recall decay. Figure 4 shows the delay between veterinary visit and subsequent Dogslife data entry. Illnesses that went unreported were again associated with a longer delay. The illness and vaccination plots have a remarkably similar shape indicating a similar pattern of recall among the owners for both questions.

Method of Communication

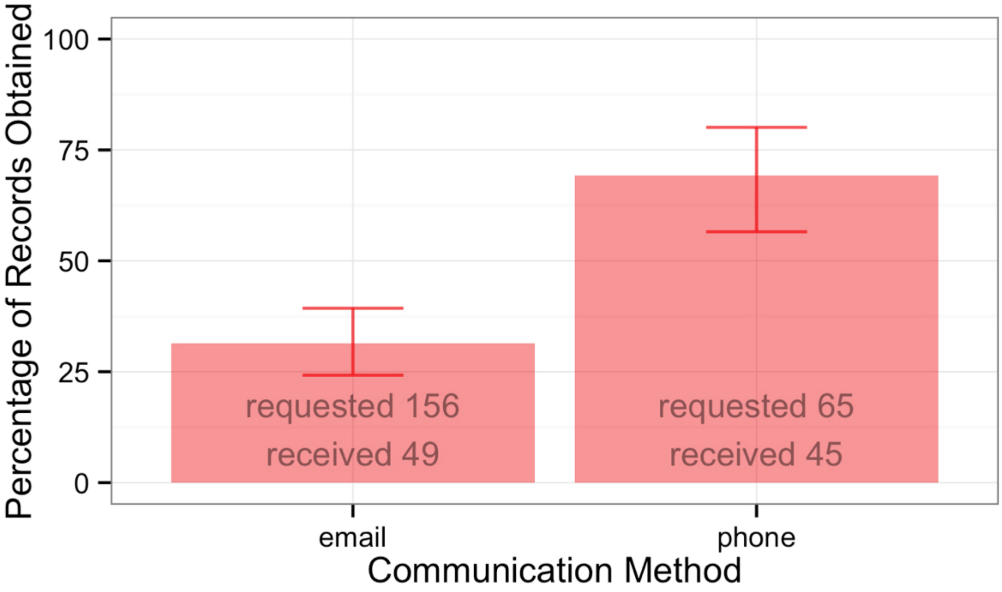
Overview

Telephone conversations were more likely to elicit a positive response to the request for data than electronic communications (Example 9). It was not determined whether this was because those who permitted telephone contact were more likely to respond per se or whether the personal interaction involved in a phone call led to a better response.

Example 9: Automated Communication Does Not Eliminate Need for Personal Interactions

Owners were contacted and then sent a form to sign to request their dogs' veterinary records. A greater proportion of records was obtained from owners contacted by phone than those contacted by email (Figure 5). A lower proportion of participants permitted phone contact (33% for telephone compared to 91% for email), which meant that email remained crucial, but the flexibility to use both methods of communication better facilitated engagement in the validation effort and project as a whole.

Figure 5. Health records obtained using telephone and email.



Ease of Data Entry

Overview

Consistent feedback from owners of older dogs during the visit process was that their questionnaire answers largely did not change. They wanted to know why they kept being asked to fill in the same answers (Example 10).

Example 10: Pre-Population May Reduce Questionnaire Fatigue

In November 2013, sleeping location and dietary answers were pre-populated in the online questionnaire for dogs over 1 year of age (Multimedia Appendix 1). As a basic data check, answers that were not amended prompted a pop-up asking whether the owner was sure that nothing has changed. Given that the Dogslife project is based on the goodwill of owners, easing data entry was thought to be invaluable in retaining the cohort.

Discussion

Principal Findings

Now that we live in the era of big data, facilitated by the collection tools the Internet provides, the thorny question of validation is readily forgotten. Questionnaire design is an iterative process of design, re-wording, piloting, and re-wording again. Nevertheless, issues are still likely to arise when a finalized questionnaire goes live and "real" data are accumulated. As such, a questionnaire would ideally continue to be assessed with respect to collection of these data as a study progresses. With many longitudinal studies, recruitment is ongoing and immediate assessment of "real" data is impossible because too few entries are available to highlight problems. A balance therefore needs to be struck between the timing of any assessment, and the strength of the conclusions that can be drawn with potentially limited data. In a longitudinal context, it might be possible to explicitly treat the initial period of data collection as an extended pilot whereby the questionnaire could be refined to reflect issues raised during validity checks. Realistically, however, time and financial pressures make such a process unlikely; it is very hard to simply throw early data away. Starting validation efforts as early as possible means that issues may be addressed before too many data are collected.

In spite of extensive testing prior to launch [8], a number of the questionnaire answer options in the questionnaire were found to be problematic (Example 1). The options that were offered affected the validity of several measures. Contributors varied in their interpretation of the household descriptor "family", and dogs that slept with another pet were inconsistently categorized as either "sleeping alone" or "sleeping with a member of the family". By waiting to start our validation until we were ready to start analysis of our hypotheses (22 months after launch), we missed the opportunity to correct such simple issues with the questionnaire early in the project. In retrospect, we should have identified these issues by starting our face-to-face validation as soon as there were enough participants and data (in our study, this was approximately 3 months after launch). For longitudinal studies, a questionnaire must be quick and simple for the participant in order to maximize retention of the cohort. A balance was struck throughout the original questionnaire between clarity and brevity, and in the examples outlined above, brevity introduced potential errors into classifications.

Changes in height and weight in dogs aged under 1 year of age (ie, dogs that were still growing) made validation of answers to these questions challenging. Delays between questionnaires being answered and owner visits also had wider ramifications for the validation process. We wrongly believed that participants would be reluctant to take part in the visit-based validation exercise, and consequently a wide time interval between last online data entry and proposed visit was chosen to create a large sampling frame. With hindsight, it would have been preferable to select from those who have given data entries within a very short time. The interval can be widened if participation rates are low. A wide timeframe impacted our ability to make useful comparisons because of changes in the household situation since data entry.

Typically, for the avoidance of the issues described above, validation should be timed closely to data entry but with contemporaneously collected data such as health records, review can take place at any time. The event count for such types of data will increase as participants age, and consequently more information can be assimilated for every record collected if assessment is delayed.

Ideally, each parameter would be validated against a reliable external data source (eg, the records of clinic visits). For certain parameters, such as illnesses that do not result in a visit to health professionals, external sources may not exist. This was a facet of health behavior that we were unable to validate directly. Alternative techniques such as daily record keeping have been used in activity studies, but these methods themselves may not be robust [11].

Many studies based on online questionnaires report minimal or no validity checks, for example [12] and [13]. Time and money are limited in any project, and finding problems with data both delays analyses and makes them more complex. Some data, for example individual human dietary information, are notoriously difficult to collect, as they vary from day to day, both in type and quantity, and people have difficulty accurately recalling details [14]. Given the relatively monomorphic diets fed to pet dogs, we anticipated that dietary information would be accurately recorded by the questionnaire. However, keeping the questionnaire relatively brief was incompatible with the unexpected complexity of canine diets. The apparent difficulty owners had with weighing and reporting daily intake was at odds with what had been perceived to be a simple measuring process. By contrast, it was anticipated that owners might have difficulty weighing a relatively large breed dog and that reported dog weights might suffer from underestimation similar to those found in human studies [15,16]. The validation procedure confounded our expectations, highlighting its utility, but also demonstrating risks involved. If data are put up for validation, the investigators must be prepared for them to be shot down.

Social desirability may lead to inaccurate responses. Self-reporting of smoking status is thought to be reliable in the general population [17] but is affected by social factors in specific groups such as pregnant women [18] and those suffering from respiratory diseases [19]. We observed misclassification in response to a relatively simple question that may be attributable to the social perception of the undesirability of smoking.

Recall decay may affect any study where events are reported in retrospect and is thought to reflect the declining strength of memory [20]. It is taken into account in order to optimize response to censuses [21], specifically studied with regard to the accuracy of self-reported drug misuse [22], and must be considered in the design of all questionnaires. To try and circumvent the problem, investigators can permit participants to enter data at any point; something that online studies are ideally suited to. Unfortunately, this does not necessarily eliminate the problem (Example 8).

If recall decay remains an issue, one can focus on events reported during a number of days prior to each data entry (a small “at-risk” period) rather than that subject’s whole record. This minimizes misclassification of subjects as controls when they should be cases. Schmidt et al [23] used this technique in a longitudinal study of the prevalence of diarrhea, and while it reduced the quantity of available data, quality was improved. Unfortunately, precise quantification of the effect of recall decay (such as determined by [24]) cannot be made with a relatively small number of health records. Using a limited “at-risk” period and, for high prevalence diseases, requesting further records to ensure accuracy of case and non-case status, will more broadly increase the power of risk factor studies.

The design of Dogslife as an Internet-based project should have minimized recall decay as owners may enter new information at their convenience and were not limited to reporting when an interviewer visits or the next questionnaire comes through the post. However, the interval between Dogslife data entries peaked at 37 days for dogs under 1 year [8] and 90 days for dogs aged over a year. Owners were returning when they received reminder emails, not when there was a recent incident to report.

The utopic goal of Internet-based longitudinal studies is that all communication can be automated and electronic to minimize costs. However, in our experience, while email was the favored method of contact for participants, telephone conversations were more likely to elicit a positive response. It seems it may not be practical to move everything into the virtual realm just yet.

Bias from loss to follow-up is a chronic and well-recognized problem in longitudinal studies [25]. Investigators need to design their questionnaire to minimize fatigue in participants who will hopefully answer it time and time again. One key decision driven by feedback from our validation process was pre-population of the questionnaire (Example 10). Mooney et al [26] addressed the issue of pre-populating factual answer fields as part of their longitudinal study of substance abuse. They showed that offering partially pre-populated questionnaires resulted in minimal time saved for those who completed them but was perceived to be “very helpful”—a positive effect on perception rather than reality. There is obviously a balance to be struck between reducing data quality and minimizing retention bias. Online studies can build checks into any pre-populated questionnaire so as to hopefully minimize any negative impact of pre-population on data quality. Where some answers do not change greatly with time, pre-population may be the lesser evil.

Conclusions

Testing the reliability and validity of longitudinal data captured by electronic means is crucial to substantiating any analyses based on those data. A plan for validation should be part of the design of a project. This builds quality into the science and ensures that funders do not expect immediate analyses. On the basis of our results, we suggest considering the following:

1. Timing is everything. Do not make validation the first stage of analysis; make it the last stage of questionnaire design. Initial validation should be performed as early as possible, particularly direct visitation as it may reveal or highlight unexpected issues not identified in pre-testing. It should also be appreciated that each question will have its own optimal time point for validation (both the time from the start of the study and the time between data entry and the validation procedure). When using contemporaneous health records, delaying allows time for data to accrue.
2. Some questions cannot be checked for reliability or validity. This should be recognized, accepted, and reported by the study.
3. If there is a delay between event occurrence and reporting, then expect recall decay. Try to determine how much impact it is having on the study. If participants have long intervals between contributions, then analyses may have to be changed to accommodate the "missing" timeframe.
4. Providing a non-automated alternative to Internet-based communication may aid in participation. This applies not only to validation efforts but also to potential detailed work with subsamples of a cohort.
5. Expect the unexpected. This is particularly relevant with regard to user-derived measurements taken by participants. Validation implies an assessment of the unknown. It can be disheartening but should not be regarded as simply a burden because it can have unexpectedly positive results. Prior concerns with regard to aspects of a questionnaire may be belied by positive validation results.

While validation has the potential to be a somewhat thankless task, our efforts generated positive feedback from our participants. The process also highlighted some issues that might have been missed if face-to-face visits had not been conducted. Our validation work not only means that we understand our data, but we believe it will help reduce future loss to follow-up and minimize selection bias.

Acknowledgments

The authors gratefully acknowledge the Dogslife contributors. The Project was initially funded by the Kennel Club Charitable Trust and has ongoing support from the Roslin Foundation. CP is funded by a BBSRC PhD scholarship.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Elements of the Dogslife questionnaire with the original response options and the modified response options (in bold font) following the validation exercise (the Dogslife questionnaire is copyrighted to the University of Edinburgh).

[PDF File (Adobe PDF File), 69KB - jmir_v17i4e96_app1.pdf]

Multimedia Appendix 2

Validation methodology.

[PDF File (Adobe PDF File), 58KB - jmir_v17i4e96_app2.pdf]

References

1. Kaaks R, Riboli E. Validation and calibration of dietary intake measurements in the EPIC project: methodological considerations. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26 Suppl 1:S15-S25 [FREE Full text] [Medline: 9126530]
2. Sieri S, Agudo A, Kesse E, Klipstein-Grobusch K, San-José B, Welch AA, et al. Patterns of alcohol consumption in 10 European countries participating in the European Prospective Investigation into Cancer and Nutrition (EPIC) project. Public Health Nutr 2002 Dec;5(6B):1287-1296. [doi: 10.1079/PHN2002405] [Medline: 12639233]
3. Leffondré K, Abrahamowicz M, Siemiatycki J, Rachet B. Modeling smoking history: a comparison of different approaches. Am J Epidemiol 2002 Nov 1;156(9):813-823 [FREE Full text] [Medline: 12396999]
4. Smith BH, Penny KI, Purves AM, Munro C, Wilson B, Grimshaw J, et al. The Chronic Pain Grade questionnaire: validation and reliability in postal research. Pain 1997 Jun;71(2):141-147. [Medline: 9211475]
5. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Appl Psychol Meas 1977;1(3):385-401.
6. Thrusfield M. Data collection and management. In: Veterinary Epidemiology. Oxford: Blackwell Science Ltd; 2005:189-213.
7. DeVellis RF. Scale development: theory and applications. 3rd ed. London: SAGE; 2012.

8. Clements DN, Handel IG, Rose E, Querry D, Pugh CA, Ollier WER, de Bronsvoort B Mark C. Dogslife: a web-based longitudinal study of Labrador Retriever health in the UK. BMC Vet Res 2013;9:13 [FREE Full text] [doi: 10.1186/1746-6148-9-13] [Medline: 23332044]
9. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989 Mar;45(1):255-268. [Medline: 2720055]
10. McBride GB. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. Hamilton, New Zealand: National Institute of Water & Atmospheric Research Ltd; 2005. URL: http://www.medcalc.org/download/pdf/McBride2005.pdf [accessed 2015-04-10] [WebCite Cache ID 6XggaZlJs]
11. Bringolf-Isler B, Mäder U, Ruch N, Kriemler S, Grize L, Braun-Fahrländer C. Measuring and validating physical activity and sedentary behavior comparing a parental questionnaire to accelerometer data and diaries. Pediatr Exerc Sci 2012 May;24(2):229-245. [Medline: 22728415]
12. Torous J, Chan S, Yee-Marie Tan S, Behrens J, Mathew I, Conrad E, et al. Patient Smartphone Ownership and Interest in Mobile Apps to Monitor Symptoms of Mental Health Conditions: A Survey in Four Geographically Distinct Psychiatric Clinics. JMIR Mental Health 2014 Dec 23;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004]
13. Wetterlin FM, Mar MY, Neilson EK, Werker GR, Krausz M. eMental health experiences and expectations: a survey of youths' Web-based resource preferences in Canada. J Med Internet Res 2014;16(12):e293 [FREE Full text] [doi: 10.2196/jmir.3526] [Medline: 25519847]
14. Karvetti R, Knuts L. Validity of the 24-hour dietary recall. Journal of the American Dietetic Association 1985;85(11):1437-1442.
15. Lassale C, Péneau S, Touvier M, Julia C, Galan P, Hercberg S, et al. Validity of web-based self-reported weight and height: results of the Nutrinet-Santé study. J Med Internet Res 2013;15(8):e152 [FREE Full text] [doi: 10.2196/jmir.2575] [Medline: 23928492]
16. Connor GS, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. Obes Rev 2007 Jul;8(4):307-326. [doi: 10.1111/j.1467-789X.2007.00347.x] [Medline: 17578381]
17. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. Am J Public Health 1994 Jul;84(7):1086-1093. [Medline: 8017530]
18. Russell T, Crawford M, Woodby L. Measurements for active cigarette smoke exposure in prevalence and cessation studies: why simply asking pregnant women isn't enough. Nicotine Tob Res 2004 Apr;6 Suppl 2:S141-S151. [doi: 10.1080/14622200410001669141] [Medline: 15203817]
19. Lewis SJ, Cherry NM, Mc LR, Barber PV, Wilde K, Povey AC. Cotinine levels and self-reported smoking status in patients attending a bronchoscopy clinic. Biomarkers 2003;8(3-4):218-228. [doi: 10.1080/1354750031000120125] [Medline: 12944174]
20. Hinrichs J. A two-process memory-strength theory for judgement of recency. Psychol Rev 1970;77(3):223-233.
21. Neter J, Waksberg J. Three-month recall loss and reference period effects. In: Response errors in collection of expenditures data by house interviews: An experimental study. Washington, DC: US Department of Commerce, Bureau of the Census; 1965:51-65.
22. Shillington AM, Reed MB, Clapp JD, Woodruff SI. Testing the length of time theory of recall decay: examining substance use report stability with 10 years of national longitudinal survey of youth data. Subst Use Misuse 2011;46(9):1105-1112 [FREE Full text] [doi: 10.3109/10826084.2010.548436] [Medline: 21406007]
23. Schmidt WP, Luby SP, Genser B, Barreto ML, Clasen T. Estimating the longitudinal prevalence of diarrhea and other episodic diseases: continuous versus intermittent surveillance. Epidemiology 2007 Sep;18(5):537-543. [doi: 10.1097/EDE.0b013e318093f3ce] [Medline: 17603390]
24. Feikin DR, Audi A, Olack B, Bigogo GM, Polyak C, Burke H, et al. Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. Int J Epidemiol 2010 Apr;39(2):450-458 [FREE Full text] [doi: 10.1093/ije/dyp374] [Medline: 20089695]
25. Rothman KJ, Greenland S. Cohort Studies. In: Modern Epidemiology 2nd ed. Philadelphia: Lippincott-Raven; 1998:79-92.
26. Mooney G, Krakowiecki M, Trunzo D. Does Prefilling Responses on a Longitudinal Establishment Survey Stem Sample Attrition? In: Am Assoc Public Opin Res. 2009 Presented at: Survey Research Methods Section, American Statistical Association conference; 2009; Washington, DC p. 5917-5924.

Edited by G Eysenbach; submitted 09.05.14; peer-reviewed by M van Gelder, T Al Baghal; comments to author 17.11.14; revised version received 06.01.15; accepted 03.02.15; published 16.04.15.

Please cite as:
Pugh CA, Summers KM, Bronsvoort BMC, Handel IG, Clements DN
Validity of Internet-Based Longitudinal Study Data: The Elephant in the Virtual Room
J Med Internet Res 2015;17(4):e96
URL: http://www.jmir.org/2015/4/e96/
doi:10.2196/jmir.3530
PMID:25887101

Original Paper

A New Method for Assessing Content Validity in Model-Based Creation and Iteration of eHealth Interventions

Nancy Kassam-Adams[1,2], PhD; Meghan L Marsac[1,2], PhD; Kristen L Kohser[1], MSW; Justin A Kenardy[3], PhD; Sonja March[4], PhD; Flaura K Winston[1,2], MD, PhD

[1]Center for Injury Research and Prevention, Children's Hospital of Philadelphia, Philadelphia, PA, United States
[2]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[3]Centre of National Research on Disability and Rehabilitation Medicine and School of Psychology, University of Queensland, Brisbane, Australia
[4]University of Southern Queensland, Springfield, Australia

Corresponding Author:
Nancy Kassam-Adams, PhD
Center for Injury Research and Prevention
Children's Hospital of Philadelphia
3535 Market St, Suite 1150
Philadelphia, PA, 19104
United States
Phone: 1 2155903118
Fax: 1 2155905425
Email: nlkaphd@mail.med.upenn.edu

Abstract

Background: The advent of eHealth interventions to address psychological concerns and health behaviors has created new opportunities, including the ability to optimize the effectiveness of intervention activities and then deliver these activities consistently to a large number of individuals in need. Given that eHealth interventions grounded in a well-delineated theoretical model for change are more likely to be effective and that eHealth interventions can be costly to develop, assuring the match of final intervention content and activities to the underlying model is a key step. We propose to apply the concept of “content validity” as a crucial checkpoint to evaluate the extent to which proposed intervention activities in an eHealth intervention program are valid (eg, relevant and likely to be effective) for the specific mechanism of change that each is intended to target and the intended target population for the intervention.

Objective: The aims of this paper are to define content validity as it applies to model-based eHealth intervention development, to present a feasible method for assessing content validity in this context, and to describe the implementation of this new method during the development of a Web-based intervention for children.

Methods: We designed a practical 5-step method for assessing content validity in eHealth interventions that includes defining key intervention targets, delineating intervention activity-target pairings, identifying experts and using a survey tool to gather expert ratings of the relevance of each activity to its intended target, its likely effectiveness in achieving the intended target, and its appropriateness with a specific intended audience, and then using quantitative and qualitative results to identify intervention activities that may need modification. We applied this method during our development of the Coping Coach Web-based intervention for school-age children.

Results: In the evaluation of Coping Coach content validity, 15 experts from five countries rated each of 15 intervention activity-target pairings. Based on quantitative indices, content validity was excellent for relevance and good for likely effectiveness and age-appropriateness. Two intervention activities had item-level indicators that suggested the need for further review and potential revision by the development team.

Conclusions: This project demonstrated that assessment of content validity can be straightforward and feasible to implement and that results of this assessment provide useful information for ongoing development and iterations of new eHealth interventions, complementing other sources of information (eg, user feedback, effectiveness evaluations). This approach can be utilized at one or more points during the development process to guide ongoing optimization of eHealth interventions.

(J Med Internet Res 2015;17(4):e95) doi:10.2196/jmir.3811

KEYWORDS

telemedicine; methods; stress disorders, post-traumatic; child; secondary prevention

Introduction

Overview

The advent of eHealth interventions to address psychological concerns and health behaviors has created new opportunities and new challenges. Some eHealth interventions are adaptations of established face-to-face interventions; many are created from scratch as electronically delivered interventions. In either case, they provide the ability to optimize the effectiveness of intervention activities and then deliver these activities consistently to a large number of individuals in need.

Ideally, development of any intervention (whether the intervention is delivered electronically or in-person) begins with a clearly delineated program theory or model of change that is grounded in empirical evidence and clinical experience [1,2]. In such a model, intervention activities target specific mechanisms (psychological or behavioral processes) in order to produce desired modifications in health or behavioral outcomes. The use of a theoretical model to guide development of an eHealth intervention appears to be associated with effectiveness. A recent meta-analysis of 85 studies of eHealth interventions for health behavior change found that interventions that made greater use of theory (ie, linking theoretical constructs to intervention techniques) had larger effect sizes [3].

Definition of Content Validity and Adaptation for eHealth Interventions

The concept of content validity originates in the arena of psychological and educational instrument development. Content validity of an assessment instrument is one aspect of construct validity [4,5] and has been defined as “the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose”, page 238 [5]. In this definition, the relevant elements of an instrument may vary based on the method and purpose of the assessment and include item wording as well as the way in which stimuli are presented, how instructions are given, and which situations are sampled [5]. Content validity is conditional rather than an inherent trait of an assessment instrument; it is assessed with regard to a particular purpose or aim of assessment, and a particular targeted population [5]. Quantitative and qualitative indicators derived from expert review of an instrument’s content validity can be useful in identifying missteps and honing content during the development phase of an assessment instrument [5,6].

Our definition of content validity for eHealth builds upon these well-established attributes of content validity in instrument development. We define the content validity of an eHealth intervention as the extent to which its component intervention activities are relevant to the underlying construct (ie, program theory) and likely to be effective in achieving a particular intervention purpose in a specific intended population. We therefore suggest three core dimensions for expert review: relevance, likely effectiveness, and appropriateness for a specific audience. The first dimension, *relevance*, is the extent to which an intervention activity is pertinent to its intended intervention target as defined in the program theory or model of change, that is, “Is this arrow aimed in the right direction?” The second dimension, *likely effectiveness*, is the extent to which evidence, theory, and expert judgment would suggest that this specific activity would successfully modify the intended intervention target, that is, “Is this arrow likely to hit the target?” The third dimension is the extent to which the activity is *appropriate for a specific intended audience*, which may be defined by age, culture, or other factors.

Rationale for Considering Content Validity of eHealth Interventions

There are several compelling reasons to attend carefully to content validity in the development of eHealth interventions. When intervention content is developed based on a clearly delineated program theory and model of change, not only is the eHealth intervention more likely to be effective, its use and evaluation also advances understanding of the psychological or behavioral processes in which one is trying to intervene [7,8]. However, after the developer of an eHealth intervention has articulated a model of change/program theory, the next steps are fraught with challenges, including a multitude of choices in the design and delivery of intervention activities. A formal assessment of content validity can be a key checkpoint in the design of actual intervention activities to ensure that these activities and processes match the underlying program theory and change model that they are intended to operationalize. Electronically delivered intervention programs (to an even greater extent than manualized in-person interventions) deliver a set of pre-determined and highly observable activities, thereby facilitating review of specific activities during the development process. Results of content validity assessment can be used to hone an eHealth intervention during its development or to better understand unexpected variations in the performance of an existing intervention. Given the expense of developing eHealth interventions, assessing content validity early in the development process (eg, at the storyboard stage) could be cost-effective by increasing the likelihood that costly further development will lead to an effective intervention.

Undoubtedly, most intervention developers strive to achieve this sort of validity and informally assess the extent to which they have succeeded. However, to our knowledge no systematic process for assessing content validity of eHealth interventions has been proposed. Thus, the aims of this paper are to define content validity as it applies to model-based eHealth intervention development, present a feasible method for assessing content validity in this context, and describe the implementation of this method during the development of an intervention. Based on the definition of content validity for eHealth proposed above, we designed a practical method for assessing the content validity of an eHealth intervention, and applied this methodology during our development of the Coping Coach Web-based intervention for school-age children. Formal review by external experts allowed us to evaluate the extent to which intervention activities matched the model of change and program theory upon which we based our intervention development.

Methods

Procedures for Expert Review of Content Validity

We propose a straightforward, systematic approach to obtaining expert review of the content validity of an eHealth intervention. This approach assumes that the eHealth intervention in question has been created based on an explicit program theory or model of change or that an appropriate theory/model can be applied (even retrospectively) to the existing intervention content. Our approach consists of the following steps:

- Step 1 is to specify key intervention targets that this e-Health intervention is intended to address.
- Step 2 is to delineate specific activity-target pairings, by defining discrete intervention activities that address each target. These may not be one-to-one relationships; a target may be addressed by more than one activity or vice versa. In this context, an intervention activity is defined as a meaningful set of user actions or experiences that can be clearly linked to one or more targets. It is important to be able to describe each activity so that expert reviewers understand exactly which intervention content is included.
- Step 3 is to populate the Content Validity Survey Tool with each intervention activity-target pairing. The Survey Tool includes scales for relevance, likely effectiveness, and appropriateness for a specific intended audience (see Multimedia Appendix 1).
- Step 4 is to recruit experts who were not involved in the development of the eHealth intervention and gather survey data using the Content Validity Survey Tool. Expert reviewers should each have relevant content knowledge; disciplinary and geographic diversity across the set of reviewers can provide useful balance [5,9]. The ideal number of reviewers has been suggested as 8-12 for an initial stage review of content validity, and 3-5 (who may be a subset of the original group) for a secondary or follow-up review [6]. Expert reviewers must be provided with access to the best current version of the eHealth intervention (eg, storyboard, text script, online access to the intervention) and asked to complete an online or emailed copy of the Content Validity Survey Tool created in Step 3.
- Step 5 is to analyze results and refine the intervention as needed. Results should be analyzed quantitatively (eg, via calculation of content validity indices) and qualitatively (eg, via examination of narrative comments from expert raters). Use these findings to identify potentially problematic activities and to hone the intervention as needed. Depending on the stage of development of the intervention, the development team may elect to remove or revise potentially suboptimal intervention activities immediately or in a future iteration of the intervention.

Based on prior literature regarding content validity in instrument development, we propose both item-level and scale-level content validity indices (I-CVI and S-CVI/AV) as quantitative indicators of acceptable content validity [6]. For eHealth interventions, an “item” is an intervention activity-target pairing, and the “scale” is a set of activity-target pairs. The I-CVI is the proportion (0.0 to 1.0) of expert reviewers who rate an item as 3 or 4 on a 4-point scale. The S-CVI/AV is the average of all I-CVIs for a set of items [6]. Polit et al [6] proposed standards for content validity based on a review of the literature and examination of the quantitative properties of alternate content validity indices. We propose to adopt these standards for eHealth, such that a set of eHealth intervention activities can be said to have excellent content validity if all I-CVIs are at least .78 and the S-CVI/AV is at least .90.

Utilizing the results of content validity assessment to hone and improve an eHealth intervention will always involve both quantitative indicators and the considered judgment of the development team. For example, if a quantitative indicator such as the I-CVI indicates problems with a specific activity-target pair, the next steps for the development team depend on the nature of the problem identified. If an activity is rated as not relevant to its intended target, the development team may consider removing it or undertaking a major revision. On the other hand, if an activity is rated as relevant but not likely to be effective, the development team should consider whether there is a way to alter or enhance the activity to increase its likely effectiveness. Narrative comments provided by expert reviewers as part of the Content Validity Survey can be helpful, and follow-up interviews to elicit additional details about specific concerns may be useful. Developers may also need to take into account whether an activity is rated as relevant/likely effective for some, but not all, of its intended targets.

Application of this Method to the Coping Coach Intervention

Description of Coping Coach

Coping Coach is an eHealth intervention designed to prevent or reduce posttraumatic stress and associated negative impacts on health-related quality of life in young people aged 8-12 years old who have experienced different types of acute, single-incident traumatic events [10]. The Coping Coach intervention is structured as an interactive game with a storyline. Intervention activities include skills practice and interactions with game characters as the child user progresses through three levels of the game. The program theory that underlies our development of the Coping Coach intervention is presented in detail elsewhere [10,11] and described briefly here. Grounded in the empirical literature on posttraumatic stress etiology [12-14], we first identified four proximal goals for users of the Coping Coach intervention: (1) identify emotional reactions after trauma, (2) build cognitive re-appraisal skills, (3) reduce use of avoidance coping strategies, and (4) increase social support seeking. For each of these goals, we drew from the empirical literature on intervention for posttraumatic stress and anxiety in children [15-17] to delineate more specific actionable intervention targets and then worked closely with a Web/game developer team to craft intervention activities to address these targets.

Application of 5-Step Content Validity Process to Coping Coach

Step 1 (specify key intervention targets) was integrated throughout the development process. As described above, we developed the intervention based on a program theory in which we identified 13 specific intervention targets (2-5 targets for each of the four proximal goals) to address key mechanisms for prevention of posttraumatic stress.

In Step 2 (delineate activity-target pairings), we delineated 11 discrete intervention activities; each activity addressed one or more of the 13 intervention targets, resulting in a total of 15 activity-target pairs. Table 1 presents each intervention activity with the target(s) it was intended to address. For the Coping Coach intervention, the process of delineating activity-target pairings was straightforward, as our intervention development process began with a careful definition of intervention targets, and each activity was designed to address one or more of these targets. When an intervention has not been developed explicitly in this manner, Step 2 may require additional effort such as mapping activity-target pairings via a consensus process among key members of the development team and/or evaluating interrater reliability in matching activities to targets.

For Step 3 (populate survey tool), we created a Content Validity Survey Tool with three ratings for each of the 15 activity-target pairs. For each activity-target pairing, experts rated the intervention activity's relevance, likely effectiveness, and age-appropriateness, using a 5-point Likert-type scale (0-4), as follows: (1) *Relevance* (the extent to which this specific intervention activity is pertinent to the intended intervention target) with 0 defined as "Irrelevant/Extraneous to this target" and 4 defined as "Central/Key/Essential to this target"; (2) *Effectiveness* (likelihood that this specific activity will successfully modify the intended intervention target), with 0 defined as "Not likely to be effective" and 4 defined as "Very likely to be effective"; and (3) *Appropriateness for intended audience*, which, in the case of Coping Coach, was defined as age-appropriateness (extent to which the language, content, and nature of activities was clear, easy to understand, and developmentally appropriate for children age 8-12 years), with 0 defined as "Inappropriate/Unsuitable for 8-12" and 4 defined as "Language/nature of activities appropriate for 8-12". The survey form included screen shots from the intervention to help orient expert reviewers to the activity they were rating. A copy of the Content Validity Survey Tool template is available in Multimedia Appendix 1.

For Step 4 (recruit experts and gather survey data), an international set of experts was invited to participate in rating the Coping Coach intervention. Experts were selected based on their knowledge and expertise regarding children's coping and adjustment after potentially traumatic events, traumatic stress prevention, culturally sensitive child interventions, or development of Web-based interventions. We provided each expert with a username and password to access the Coping Coach intervention online and encouraged them to play through the entire game at least once as a child user would. We also provided a full text transcript of all intervention elements and activities. The Content Validity Survey Tool was provided as a Word document and sent to experts via email; experts completed their ratings within this document and returned the document via email. Expert reviewers were asked to complete the Content Validity Survey Tool and to provide additional comments on any specific activity or on the intervention as a whole. Expert review of an intervention does not constitute human subjects research, and thus no Institutional Review Board or ethics board approval process was relevant or required.

For Step 5 (analyze results and hone intervention), we first calculated the I-CVI for each activity-target pair on each dimension; the I-CVI is the proportion of reviewers who gave a rating of 3 or 4 on the 5-point scale (0-1-2-3-4) used in this version of the Survey Tool. We then calculated the S-CVI/AV for each dimension (relevance, likely effectiveness, and age-appropriateness) as the average of all I-CVIs for that dimension. In this case the I-CVI is a slightly more conservative indicator of expert consensus than described by Polit et al, because the survey tool utilized for ratings of Coping Coach used a 5-point scale rather than a 4-point scale. We also examined additional narrative comments from the expert reviewers. These data, in conjunction with feedback from child users and their parents [10] and the results of a pilot randomized trial [11], are now being used to hone and improve the next iteration of the Coping Coach intervention.

Results

Overview

A total of 15 experts (from the United States, Australia, United Kingdom, Netherlands, and Switzerland) were invited to participate via email; all 15 agreed to participate and provided ratings. All experts were independent, that is, not involved in the development of the intervention. Each expert was an active clinical researcher (12 psychologists and 3 psychiatrists) with at least 5 years of experience in this field and relevant content expertise. Thirteen experts provided ratings within approximately 2 months as requested; 2 experts required additional time due to other commitments but eventually provided ratings. There were very few incomplete ratings (only 2 of 225 ratings for relevance, 3 of 225 ratings for effectiveness, and 5 of 225 ratings for age-appropriateness were missing).

Quantitative Indicators of Content Validity

Table 1 shows the I-CVI for each intervention activity/target pairing, based on expert ratings from the Content Validity Survey Tool.

Table 1. Intervention activities and intervention targets for each of four proximal goals of the intervention, with item-level Content Validity Index for each activity-target pairing.

Intervention activity	Intended intervention target(s) for this activity	Item-level Content Validity Index		
		Relevance of activity to target	Likely effectiveness of activity	Age appropriateness of activity
A. Proximal goal: Identify emotional reactions (EM)				
1. Player creates faces by manipulating the eyebrows and mouth to match specified feelings in Face-O-Matic Machine.	EM1: Child will identify and name basic feelings/ emotions.	.80	.60	.73
2. Player helps the townspeople identify how they were feeling at the time of a potentially traumatic event, and how they are feeling now.	EM2: Child will identify feelings/ emotions associated with a potentially traumatic experience, and how these feelings may change over time.	.93	.93	.93
3. Player identifies own feelings with the help of the townspeople.	EM3: Child will identify their own feelings associated with a potentially traumatic experience, and any changes in these feelings over time.	.93	.86	1.0
B. Proximal goal: Build cognitive re-appraisal skills (CR)				
4. Player watches/ listens to conversation between General Malaise and the Coping Coach about Think→ Feel→ Act.	CR1: Child will recognize connections between thoughts (appraisals), feelings, and behavior.	1.0	.73	.73
	CR2: Child will recognize helpful/ unhelpful trauma-related thoughts/ appraisals and see appraisals as something that can be modified.	.93	.73	.67
5. Player helps Jack and Jayla understand their thoughts and feelings and then helps Jack/Jayla change unhelpful thoughts in order help them to feel better.	CR3: Child will identify helpful/ unhelpful trauma-related appraisals.	1.0	.87	.87
	CR4: Child will use cognitive restructuring to modify unhelpful appraisals.	1.0	.80	.79
6. Player identifies own helpful/unhelpful thoughts by selecting whether statements are “like me” or “not like me”, and player’s helpful thoughts guide the airship upwards.	CR5: Child will identify their own helpful and unhelpful thoughts/ appraisals and apply cognitive restructuring to modify own unhelpful appraisals.	.93	.80	.86
C. Proximal goal: Reduce use of avoidance coping strategies (AV)				
7. Coping Coach describes Avoidance and Approach strategies. Player helps townspeople identify pros/cons of avoidance, sees 2 people modeling approach strategies, and helps 2 people replace avoidance with approach strategies.	AV1: Child will identify pros/ cons of avoidance and approach strategies for trauma-related fears/situations.	1.0	.93	.87
	AV2: Child will approach trauma-related fears situations safely and minimize reliance on avoidant coping strategies.	1.0	.87	.87
8. Sorting activity – player fixes weather machine by correctly identifying pros/cons and impact of using avoidance or approach strategies for trauma-related fears/ situations.	AV1: Child will identify pros/ cons of avoidance and approach strategies for trauma-related fears/ situations.	.87	.93	.80
	AV2: Child will approach trauma-related fears situations safely and minimize reliance on avoidant coping strategies.	.80	.80	.79
D. Proximal goal: Increase social support seeking (SS)				
9. Player gives and receives help to/ from the townspeople and General Malaise.	SS1: Child will ask for help and build support network by providing help to others.	1.0	.80	1.0
10. Player completes logbook pages to identify “People Who Can Help Me” and “Ways That People Can Help Me”.	SS2: Child will identify members of their support network and what type of support network can offer.	1.0	.93	.93

Intervention activity	Intended intervention target(s) for this activity	Item-level Content Validity Index		
		Relevance of activity to target	Likely effectiveness of activity	Age appropriateness of activity
11. Player collects coins scattered throughout the worlds – six of these coins have tips for social support seeking.	SS3: Child will increase strategies for asking for help/ social support.	.86	.79	.86

The S-CVI/AV for ratings across all activities was excellent for relevance (.94), although not quite at this standard for likely effectiveness (.82) or age-appropriateness (.85). Examining the quantitative indicators at a more granular level, the I-CVIs for likely effectiveness and age-appropriateness were excellent (≥.78) for nearly all activity-target pairings. However, we identified two activities (encompassing three activity-target pairs) with I-CVIs below .78. Intervention activity 1 (Face-O-Matic Machine activity) and intervention activity 4 (conversation between Coping Coach and General Malaise character), had excellent ratings for relevance but had I-CVIs of .60 to .73 for likely effectiveness or age-appropriateness. Reviewers also provided narrative comments to explain their concerns and/or suggest alternate approaches. Based on these I-CVIs and review comments, these two activities are under review to determine whether they should be retained, removed, or modified in the next iteration of the intervention.

Narrative Comments From Expert Reviewers

Beyond quantitative ratings, for many intervention activities the reviewers' narrative comments were helpful in understanding both strengths and potential gaps in this iteration of the Coping Coach intervention. Reviewers commented on likely mechanisms of action, for example, "One of strongest sections, teaches link between thoughts, feelings, and actions well, and good in identifying concrete thoughts" and "Interactive nature of the exercise and the fact that it doesn't 'sugar coat' that there are some positives to avoidance is useful as it makes it realistic for kids". Reviewers also highlighted ways to extend or improve current intervention activities to better achieve key targets, for example, "Perhaps also discussing what a child's behavioral reactions may be when sad, angry, worried, etc (for example, crying, stamping feet, churning stomach) may offer them more of a chance to identify their feelings".

Discussion

Principal Findings

This project demonstrated that assessment of content validity was straightforward and feasible to implement and that results of this assessment can provide useful information for ongoing development and iterations of new eHealth interventions. Expert ratings on the Content Validity Survey Tool demonstrated variability, suggesting that response options were appropriately scaled and anchored to capture useful gradations in expert judgment about the content validity of specific intervention activities.

Especially for components believed to be key to intervention outcomes, assessment of content validity could reduce the number of iterations needed to produce an effective eHealth intervention. The clear articulation of a model of change and program theory, and content validity assessment to ensure that intervention activities match their intended targets, may be especially important in the development of eHealth interventions that are created from scratch, that is, that are not Web-based adaptations of an existing well-established face-to-face intervention [10,18]. However, content validity assessment may also be beneficial to ensure that well-established face-to-face interventions are successfully translated for Web-based/digital delivery. The translation of in-person treatment components (eg, exposure to address anxiety symptoms) is not always straightforward.

Expert ratings of content validity can be an important complement to other sources of information. Depending on the point in the development cycle when content validity is assessed, an intervention development team may need to weigh information regarding suboptimal content validity of specific activities in the context of user feedback, effectiveness evaluations, and other contextual considerations to determine appropriate action. One possibility is immediate refinement or removal of potentially problematic activities; another option is ongoing monitoring of these activities in terms of user engagement or effectiveness. In the case of Coping Coach, expert review affirmed the relevance of all rated activities but identified potential gaps in the likely effectiveness and age-appropriateness of two intervention activities. This information will be utilized in conjunction with user feedback and results of a pilot randomized trial to make decisions about optimizing Coping Coach intervention activities.

Implementation of this method revealed a number of lessons regarding the process and timing of assessing content validity. Regarding process, we learned that thoughtful judgment by the development team is required to define intervention activities at an appropriately granular level, that is, with just enough specificity for meaningful evaluation by expert reviewers. Thoughtful judgment is also required to identify which activity-target pairs merit assessment of content validity. In the case of Coping Coach, we chose to structure the survey with each intervention target paired with the one or two activities that addressed that target most directly. However, some intervention targets are addressed at least indirectly by additional activities and a longer survey form could have asked expert raters to assess all such activity-target pairings. One potential outcome of Steps 1 and 2, or of feedback received during expert review (Step 4), is that intervention developers may realize that they have not adequately specified intervention targets or the intended match between intervention activities and intended targets. If this occurs, it can be seen as an important reminder for the development team, aided by expert consultation if necessary, to revisit and clarify the program theory and model of change that underlie the eHealth intervention. Clarity in this regard is likely to be helpful not only in intervention design, but also in promoting effectiveness of the intervention.

Regarding timing, we sought expert review of content validity at a point in the development process when we had already created a functional online intervention, piloted this intervention with child users, and initiated a pilot randomized trial. The advantage of this timing is that experts saw a fully developed version of the intervention activities and could fully grasp our intended design. There would be different advantages to seeking expert review earlier (with storyboards or functional prototypes) or at multiple points in the development process, namely the ability to iteratively revise an earlier draft of the intervention based on content validity assessment.

Limitations

There are several limitations of this project that suggest future research directions. First, we suggest the application of quantitative indicators (the I-CVI and S-CVI/AV), and threshold levels for those indicators, which are based in content validity research for the development of psychological measures. While we believe this is a reasonable place to start, additional research is needed to document the range of I-CVI and S-CI/AV results in the development of a variety of types of eHealth intervention activities, and the relationship of these ratings to improved performance of eHealth interventions. Such research would also help to assess the reliability and validity of the proposed Content Validity Survey Tool itself, as this was beyond the scope of the current study. It is important to note that we implemented this content validity approach for an eHealth program with a “tunnel” design, in which all users are required to participate in all activities. Because eHealth interventions vary in the extent to which every user is directed to participate in the same set of activities, assessment of content validity for more complex branching structures, or for interventions that allow free exploration of a set of activities, may require some adaptation of our method. The proposed method for content validity targets three domains: relevance, likely effectiveness, and appropriateness for a specific audience. As the tool becomes more widely used, a need might be recognized for additional domains. Finally, it is essential to remember that content validity as rated by experts is no guarantee of the effectiveness of a set of intervention activities. Researchers and intervention developers should not substitute content validity assessment for rigorous assessment of intervention effectiveness.

Conclusions

Content validity assessment can be a helpful checkpoint in the process of developing or improving an eHealth intervention. Our team created and implemented a straightforward method and Content Validity Survey Tool that provided useful information regarding the match of intervention activities to underlying program theory. This approach could be appropriately utilized at multiple points during the development process to guide ongoing optimization of eHealth interventions.

Acknowledgments

This work was funded by grant R21HD069832 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development in the United States. We gratefully acknowledge the efforts of the expert reviewers who participated in this project: Melissa Alderfer, Lamia Barakat, Steven Berkowitz, Melissa Brymer, Cindy Buchanan, Judith Cohen, Catherine Cox, Douglas Delahanty, Alexandra De Young, Markus Landolt, Richard Meiser-Steadman, Joanne Mouthaan, Reginald Nixon, Miranda Olff, and Josef Ruzek.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Content validity survey tool.

[PDF File (Adobe PDF File), 343KB - jmir_v17i4e95_app1.pdf]

References

1. Rogers P, Petrosino A, Huebner T, Hacsi T. Program theory evaluation: Practice, promise, and problems. New Directions for Evaluation 2000;87:4-13.
2. Winston FK, Jacobsohn L. A practical approach for applying best practices in behavioural interventions to injury prevention. Inj Prev 2010 Apr;16(2):107-112 [FREE Full text] [doi: 10.1136/ip.2009.021972] [Medline: 20363817]
3. 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]
4. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955 Jul;52(4):281-302. [Medline: 13245896]
5. Haynes S, Richard D, Kubany E. Content validity in psychological assessment: A functional approach to concepts and methods. Psychol Assess 1995;7(3):238-247.
6. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007 Aug;30(4):459-467. [doi: 10.1002/nur.20199] [Medline: 17654487]

7. 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]
8. Insel T. Director's Blog 27 Feb. 2014. A new approach to clinical trials URL: http://www.nimh.nih.gov/about/director/2014/a-new-approach-to-clinical-trials.shtml [accessed 2014-08-26] [WebCite Cache ID 6S7LzgFR0]
9. Grant JS, Davis LL. Selection and use of content experts for instrument development. Res Nurs Health 1997 Jun;20(3):269-274. [Medline: 9179180]
10. Marsac ML, Winston FK, Hildenbrand A, Kohser KL, March S, Kenardy JA, et al. Systematic, theoretically-grounded development and feasibility testing of an innovative, preventive web-based game for children exposed to acute trauma Clin Pract Pediatr Psychol. Clinical Practice in Pediatric Psychology 2015;3(1):12-24.
11. Marsac ML, Kohser KL, Winston FK, Kenardy J, March S, Kassam-Adams N. Using a web-based game to prevent posttraumatic stress in children following medical events: design of a randomized controlled trial. Eur J Psychotraumatol 2013;4:21311 [FREE Full text] [doi: 10.3402/ejpt.v4i0.21311] [Medline: 23898396]
12. Alisic E, Jongmans MJ, van WF, Kleber RJ. Building child trauma theory from longitudinal studies: a meta-analysis. Clin Psychol Rev 2011 Jul;31(5):736-747. [doi: 10.1016/j.cpr.2011.03.001] [Medline: 21501581]
13. Meiser-Stedman R, Dalgleish T, Glucksman E, Yule W, Smith P. Maladaptive cognitive appraisals mediate the evolution of posttraumatic stress reactions: A 6-month follow-up of child and adolescent assault and motor vehicle accident survivors. J Abnorm Psychol 2009 Nov;118(4):778-787. [doi: 10.1037/a0016945] [Medline: 19899847]
14. Trickey D, Siddaway AP, Meiser-Stedman R, Serpell L, Field AP. A meta-analysis of risk factors for post-traumatic stress disorder in children and adolescents. Clin Psychol Rev 2012 Mar;32(2):122-138. [doi: 10.1016/j.cpr.2011.12.001] [Medline: 22245560]
15. Cohen J, Mannarino A, Deblinger E. Treating Trauma and Traumatic Grief in Children and Adolescents. New York: Guilford Press; 2006.
16. Smith P, Yule W, Perrin S, Tranah T, Dalgleish T, Clark DM. Cognitive-behavioral therapy for PTSD in children and adolescents: a preliminary randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2007 Aug;46(8):1051-1061. [doi: 10.1097/CHI.0b013e318067e288] [Medline: 17667483]
17. March S, Spence SH, Donovan CL. The efficacy of an internet-based cognitive-behavioral therapy intervention for child anxiety disorders. J Pediatr Psychol 2009 Jun;34(5):474-487 [FREE Full text] [doi: 10.1093/jpepsy/jsn099] [Medline: 18794187]
18. Kassam-Adams N, Marsac ML, Winston FK. Preventing traumatic stress after child injury: Development of a website for parents. In: Brunet A, Ashbaugh AR, Herbert CF, editors. Internet Use in the Aftermath of Trauma. Amsterdam, The Netherlands: IOS Press; 2010:157-178.

Abbreviations

I-CVI: Item-level Content Validity Index
S-CVI/AV: Scale-level Content Validity Index / Average (averaged across I-CVIs)

Edited by G Eysenbach; submitted 26.08.14; peer-reviewed by T Palermo, C Knaevelsrud; comments to author 29.10.14; revised version received 26.01.15; accepted 19.02.15; published 15.04.15.

Please cite as:
Kassam-Adams N, Marsac ML, Kohser KL, Kenardy JA, March S, Winston FK
A New Method for Assessing Content Validity in Model-Based Creation and Iteration of eHealth Interventions
J Med Internet Res 2015;17(4):e95
URL: http://www.jmir.org/2015/4/e95/
doi:10.2196/jmir.3811
PMID:25881584

Review

3D Immersive Patient Simulators and Their Impact on Learning Success: A Thematic Review

Robert Kleinert[1], MD; Roger Wahba[2], MD; De-Hua Chang[3], MD; Patrick Plum[1], MD; Arnulf H Hölscher[2], MD, PhD; Dirk L Stippel[2], MD, PhD

[1]Department of General, Visceral and Cancer Surgery, University of Cologne, Cologne, Germany
[2]Department of General, Visceral and Cancer Surgery, Cologne, Germany
[3]Department of Radiology, Cologne, Germany

Corresponding Author:
Robert Kleinert, MD
Department of General, Visceral and Cancer Surgery
University of Cologne
Kerpener 62
Cologne, 50937
Germany
Phone: 49 1774445859
Fax: 49 22147887220
Email: robert.kleinert@uk-koeln.de

Abstract

Background: Immersive patient simulators (IPSs) combine the simulation of virtual patients with a three-dimensional (3D) environment and, thus, allow an illusionary immersion into a synthetic world, similar to computer games. Playful learning in a 3D environment is motivating and allows repetitive training and internalization of medical workflows (ie, procedural knowledge) without compromising real patients. The impact of this innovative educational concept on learning success requires review of feasibility and validity.

Objective: It was the aim of this paper to conduct a survey of all immersive patient simulators currently available. In addition, we address the question of whether the use of these simulators has an impact on knowledge gain by summarizing the existing validation studies.

Methods: A systematic literature search via PubMed was performed using predefined inclusion criteria (ie, virtual worlds, focus on education of medical students, validation testing) to identify all available simulators. Validation testing was defined as the primary end point.

Results: There are currently 13 immersive patient simulators available. Of these, 9 are Web-based simulators and represent feasibility studies. None of these simulators are used routinely for student education. The workstation-based simulators are commercially driven and show a higher quality in terms of graphical quality and/or data content. Out of the studies, 1 showed a positive correlation between simulated content and real content (ie, content validity). There was a positive correlation between the outcome of simulator training and alternative training methods (ie, concordance validity), and a positive coherence between measured outcome and future professional attitude and performance (ie, predictive validity).

Conclusions: IPSs can promote learning and consolidation of procedural knowledge. The use of immersive patient simulators is still marginal, and technical and educational approaches are heterogeneous. Academic-driven IPSs could possibly enhance the content quality, improve the validity level, and make this educational concept accessible to all medical students.

(J Med Internet Res 2015;17(4):e91) doi:10.2196/jmir.3492

KEYWORDS

immersive patient simulators; Web-based learning; validity; immersion; procedural knowledge

Introduction

One key factor of clinical education is the transfer of declarative knowledge (ie, "what to do") into procedural knowledge (ie, "how to do"). It is performed most effectively in small groups accompanied by a medical teacher [1]. However, in the daily practice clinical instruction is often based on a traditional apprenticeship process and on the expectation that if a student spends enough time in a clinical environment he or she will eventually "get it". Although this education is good clinical practice and known to be successful, it is impaired by increasing workload of hospital doctors, restrictive working time directives, and changes of students' attitudes and expectations in the sense of the Generation Y [2]. Therefore, new educational strategies (eg, skills labs or mannequin simulators) are developed [3,4]. Virtual patient simulators go one step further as they allow case-based learning on personal computers. In a blended learning context they can have a positive impact on knowledge gain [5]. The available simulators vary greatly in realism and interaction grade, however it is questionable whether this factor affects learners' outcome. Technological advances have allowed virtual patient simulators, such as immersive patient simulators (IPSs), with a high level of realism and interaction grade similar to computer games (ie, serious games). IPSs must not necessarily be installed on the user's home computer (ie, workstation based). New technologies enable streaming of even complex programs via the Internet directly into the user's browser (ie, Web-based). Web-based IPSs are characterized by representation of a three-dimensional (3D) virtual environment via the browser where users can freely interact in real time with their surroundings, and thus become part of the synthetic world, individually or in virtual teams (see Figure 1).

The user can playfully immerse himself/herself into the digital environment and faces the consequences of different decisions (ie, trial and error) without putting real patients at risk. By repetitive and playful training of medical procedures, procedural knowledge can be internalized and consolidated. Therefore, IPSs potentially allow time- and location-independent learning and an effective preparation for bedside teaching. However, evaluation of the impact of IPSs on knowledge gain is a demanding task, as gain in procedural experience can hardly be objectivized and is influenced by many educational factors. For medical teachers, information about validity and usability are essential parameters for a possible implementation of such simulators in the current medical curriculum. Furthermore, development of such a simulation is time- consuming and cost-intensive and only worthwhile when there is a proven effect on knowledge gain.

It was the aim of this paper to give a thematic review of the available immersive patient simulators in virtual worlds (VWs) and to evaluate whether the use of these simulators have an impact on knowledge gain by summarizing the existing validation studies.

Figure 1. Example of a 3D virtual environment streamed via the Internet directly onto the learner's computer (ie, Web-based) where the user can freely walk around and interact with the environment—authors' project in collaboration with Clemson University, SC, USA.

Methods

A systematic search of literature via PubMed of articles from 1986 to 2014 on IPSs was performed with a focus on education of medical students. IPSs were defined as digital environments that simulate medical workflows and show the characteristics of real-time simulation with free interaction in a 3D setting. As browser technology is developing fast, not only workstation-based simulators, but also Web-based simulators were included. Simulation of one-step procedures (ie, intubation, puncturing) were excluded. Search terms included a combination of "simulation," "virtual reality," or "virtual worlds," and "education" *or* "training." Peer-reviewed publications from 1986 until 2014 were included. Eligibility assessment was performed independently by two reviewers and disagreements between reviewers were resolved by consensus with a third reviewer. Titles and abstracts of all articles were screened with regard to relevance and consequently grouped—relevant articles and articles of unknown relevance were screened in full text.

Furthermore, reference lists of relevant articles were searched for additional articles that were possibly not identified until then. Identified immersive patient simulators were classified by the field of application, technical parameters (ie, Web or workstation based, immersion grade, features), medical content (*low* ≤3 cases, *medium* = 4 to 8 cases, *high*≥9 cases), and existing validation studies.

Validation was assessed according to the consensus guidelines for validation of virtual reality surgical simulators [6]: (1) concordance (aka, face) validity refers to the degree of resemblance between simulator training and training in reality, (2) content validity refers to the degree to which simulated content covers the dimension of the construct it aims to educate, (3) construct (aka, contrast) validity describes the impact of existing knowledge on simulator performance, and (4) predictive validity describes the simulator impact on future performance applications [7,8]. Validation studies were further classified in accordance with the Cochrane Handbook for Systematic Reviews of Interventions. Studies were parameterized by methodical aspects (ie, study design, level of content), technical details of the IPS, end points, and study results. Validity of observational studies was assessed by using the methodological index for nonrandomized studies [9]. Literature review and data extraction was performed by two reviewers independently, and compared afterwards.

Results

Overview

The systematic literature search (Figure 2) identified 16,946 publications that matched the criteria.

A total of 13 publications were identified as relevant in terms of describing the use of virtual worlds in medical education (Table 1). Of these publications, 9 described IPSs that were Web-based and 4 described IPSs that were workstation based.

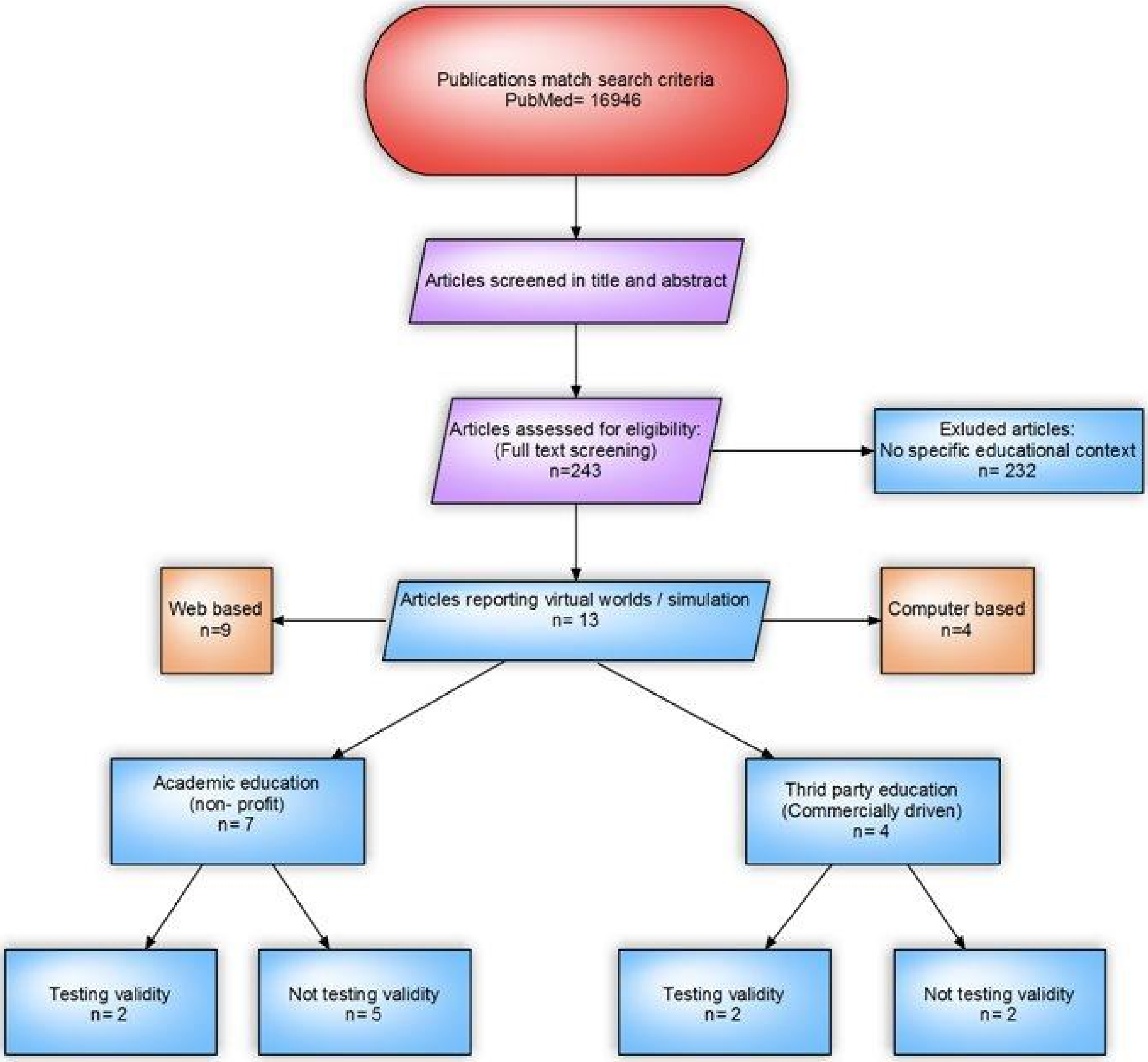
Table 1. List of available virtual patient simulators.

Virtual world	Type	Content	Immersion	Content level	Reference
VNEC	WB[a]	Neurological disorders	Medium	Medium	[10]
Play2Train	WB	Emergency medicine	Medium	Medium	[11]
MeRiTS	WB	Emergency medicine	Medium	Low	[12]
Second Health London	WB	Emergency medicine	Medium	Medium	[13]
CliniSpace	WB	Emergency medicine	Very high	High	[14]
Pulse	WB	Emergency medicine	High	High	[15]
3D Emergency Department	WB	Emergency medicine	Medium	High	[16]
Inmedea	WB	Various clinical cases	Low	High	[17]
Olive	WB	Emergency medicine	Medium	Low	[18]
Project TOUCH	CB[b]	Various clinical cases	High	Medium	[19]
Virtual Emergency Department	CB	Emergency medicine	High	High	[20]
TriageTrainer	CB	Triage	High	Low	[21]
Burn Center	CB	Burn wounds	High	Medium	[22]

[a] Web-based (WB).

[b] Computer-based (CB) (aka, workstation based).

Figure 2. Search strategy for literature on virtual patient simulators.



Web-Based Simulators

Out of 9 Web-based IPSs, 5 of them use the 3D technology of the widely used social network, Second Life (SL). All of these 5 IPSs were developed by university groups. None of these IPSs are routinely used in the education of medical students or medical staff, as they are technical feasibility projects with general medical content and focus on medical knowledge exchange [11]. However, they show the technical feasibility of providing medical education via virtual worlds [10]. A total of 3 of the Second Life-embedded virtual worlds were designed for team-based training in medical education—Second Health London is provided by the Imperial College London and was used in a pilot feasibility study including a first validation [13]. MeRiTS also offers team-based training in an SL environment, which was summarized in a feasibility study [12] without validation so far. Due to the technical specifications of Second Life, the graphical quality is not state of the art. This limitation potentially influences learners' immersion, as graphical quality and presentation have an effect on immersion grade [23].

Web-based simulators are comparable to SL simulators, however, the main difference is the technical background. Web-based simulators use a proprietary 3D engine program, which makes development of 3D worlds more resource intensive, but enables more possibilities in graphical quality and custom program design. Of the 4 remaining Web-based IPSs, 1 was already presented in 2003 and uses a custom-made 3D world. Again, the key aspect of the simulator is a feasibility study, in this case particularly with regard to distance learning [19,24]. The 3 remaining Web-based IPSs were built by commercial software companies. Of these, 2 of them show a high 3D graphical quality and, thus, a high immersion grade. They offer team- and case-based training of acute clinical cases, with the main focus on emergency medicine [14]. The high quality is reflected by the elevated pricing of these simulations—more than US $5000 for a 1-year subscription. The medical content is high and custom-made virtual cases can be added for an additional fee. The CliniSpace IPS has already been used by Stanford University for emergency procedure training in the trauma room [16,20]. The remaining IPS, Inmedea, offers low immersion as it does not use a 3D engine, but consists of freeze-frames with drawn graphics. However, the amount of medical content in this simulator is very high, and custom cases can also be added when necessary. Although it is commercially driven, there are several universities that already use this simulator in a blended learning concept [25,26], and there are already first validation studies [17].

Workstation-Based Simulators

Workstation-based IPSs need to be installed on a computer and offer high immersion with high-fidelity 3D graphics, as they are not limited by the technical handicaps of Internet broadcasting. Out of the 4 simulators, 3 of them are still undergoing feasibility studies, as they were not already validated for their effectiveness [22,27]. Although 1 of the simulators was introduced 10 years ago, it is neither routinely used in student education, nor has it been validated yet [19]. The remaining commercial VW environments by the University of Birmingham offer training in triage casualties in a simulated live exercise [21], which was part of the first validation study.

Validation

The literature search uncovered 5 articles that included validation studies, however, no study assessed all levels of validity. One of the available VW simulators validated the coherence between simulated content and real content (ie, content validity) in training of emergency procedures [13]. The correlation between the outcome of simulator training and alternative training methods was proved in 2 virtual worlds [16,20] (ie, concordance validity)—in all 3 validation studies, simulator training was comparable to alternative training in terms of outcome. The coherence between measured outcome and future performance (ie, predictive validity) was validated in 2 simulators [21,17] and showed a positive correlation. The simulators and corresponding validation studies are summarized in Table 2.

Table 2. Overview of the validity level of all currently available virtual patient simulators.

Validity type and study		Number of participants	Virtual world	Reference
Content validity				
	Cohen et al, 2013	23	Second Health London	[13]
Concordance validity				
	LeRoy Heinrichs et al, 2008	30	3D Emergency Department	[16]
	Youngblood et al, 2008	30	Virtual Emergency Department	[20]
Predictive validity				
	Funke et al, 2012	116	Inmedea	[17]
	Knight et al, 2010	91	Triage Trainer	[21]

Discussion

Principal Findings

This paper gives an overview of the available IPSs and corresponding validation studies and, therefore, summarizes the current situation in learning in virtual worlds. Previous studies revealed that using virtual patient simulators can have a positive impact on learning success [5], whereas the effect of realism grade on knowledge gain is still under discussion. IPSs as a subgroup of virtual patient simulators offer high interaction grade and realism as in computer games (ie, serious games). However, we intentionally separated the group of immersive patient simulators from serious games, as the focus of these simulators lies more on knowledge transfer than on classical game elements. The potential of IPSs lies in the internalization of diagnostic and therapeutic procedures, such as resuscitation or diagnosis patterns. Ideally, the underlying procedures are already predefined to achieve performance uniformity, similarly to standard operating procedures in the clinical daily routine [28]. It is known that immersion plays a fundamental role in virtual reality simulators, as identification with the avatar influences motivation and improves learning success [23]. However, immersion grade is influenced by many factors [29] and, therefore, hard to verify. A study from 1999 revealed that “there were no statistical differences in presence or reality judgment between a high-impact workstation and a PC workstation” [30]. However, this study was conducted before the high-fidelity graphic era and the results are not entirely transferable, as the high-impact workstation (Silicon Graphics) from 1999 was still far away from the 3D capabilities of current standard personal computers, and realism grade was comparably poor. Furthermore, the majority of the students were not used to computer worlds and, thus, less susceptible to learning with 3D worlds. Newer studies revealed that there was a positive impact of high-fidelity visual presentation on degree of immersion [31] and even on learning performance [23]. For clinical teachers, the question arises whether IPSs can support the daily routine and have an effect on students’ future performance. Assessment of new educational concepts includes different forms of validity. Content validity describes the correlation between simulated and real content. There was a positive correlation between IPS learning and the standard clinical curriculum when training students on a well-defined procedure like triage training [21] or basic procedural workflow in emergency patterns [16,20]. In times of limited time resources in the daily clinical workflow, it is desirable that the clinical curriculum be effective both for teachers and students. IPSs can facilitate students’ preparation as they allow time- and location-independent learning at an individual learning pace and with repetitions, which ensure the attainment of a similar knowledge level by the participating students. Verification of predictive validity illustrates that there is an impact of learning with IPSs on future performance, and consequently enables successful preparation for hands-on training [17,21]. However, all current IPSs are used in the blended learning context, as postprocedural review of students' performance immediately after training is known to be essential for an adequate knowledge gain [32]. Moreover, there is evidence that preexisting knowledge has a positive impact on simulator performance (ie, construct validity), although the group size of this study was low [13]. None of the articles assessed all forms of validity. As the studies are heterogeneous in medical content, the number of participants, and type of simulators, the initial question of whether the use of IPSs in the daily curriculum is beneficial in terms of learning success cannot be fully answered at this time. However, validation of single parameters revealed that IPSs can potentially support clinical teaching, although teachers must be aware of the limitations. IPSs are limited in terms of teaching declarative and procedural knowledge. Clinical education is not limited to teaching standard operating procedures, but is furthermore characterized by weighing clinical findings, evaluation of different hypotheses, and clinical experience. Therefore, IPSs are not intended to replace clinical teachers, but should support young students without relevant clinical experience. Regarding content, currently available virtual reality patient simulators range from small procedures, such as triage training, to complex procedures, such as emergency room protocols, up to teamwork with user-user interaction. The more complex the simulation, the more resources are needed. Therefore, it is not surprising that the complex simulations are commercially driven. Academic teachers can rent these simulators, making them more cost-effective, but it is questionable whether the development of a teaching method should be delegated. A highly immersive, multiuser, virtual reality simulator developed and supervised by an expert team of academic teachers would potentially allow, not only that the multiple users would train on one case (ie, team play), but also that one student would train on more than one patient at a time (ie, multitasking). These nontechnical skills can hardly be taught in reality, but are recognized as potential risk factors in high-risk environments like the emergency room or operating room [33]. It is the responsibility of universities and teaching hospitals to enable teaching methods that improve patient safety, reduce errors, and to further validate these new teaching methods.

Conclusions

Immersive patient simulators can potentially promote learning and consolidation of procedural knowledge. Web-based simulators allow time- and location-independent learning at an individual pace. The use of immersive patient simulators is still marginal, and technical and educational approaches are heterogeneous. Academic-driven IPSs could possibly enhance the content quality, improve the validity level, and make this educational concept accessible to all medical students. The development and validation of such a simulator will be the subject of our future research.

Conflicts of Interest

None declared.

References

1. Sutkin G, Wagner E, Harris I, Schiffer R. What makes a good clinical teacher in medicine? A review of the literature. Acad Med 2008 May;83(5):452-466. [doi: 10.1097/ACM.0b013e31816bee61] [Medline: 18448899]
2. Schlitzkus LL, Schenarts KD, Schenarts PJ. Is your residency program ready for Generation Y? J Surg Educ 2010;67(2):108-111. [doi: 10.1016/j.jsurg.2010.03.004] [Medline: 20656608]
3. Zendejas B, Brydges R, Wang AT, Cook DA. Patient outcomes in simulation-based medical education: a systematic review. J Gen Intern Med 2013 Aug;28(8):1078-1089 [FREE Full text] [doi: 10.1007/s11606-012-2264-5] [Medline: 23595919]
4. Gerdes B, Hassan I, Maschuw K, Schlosser K, Bartholomäus J, Neubert T, et al. [Instituting a surgical skills lab at a training hospital]. [Article in German]. Chirurg 2006 Nov;77(11):1033-1039. [doi: 10.1007/s00104-006-1212-5] [Medline: 16917754]
5. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med 2010 Oct;85(10):1589-1602. [doi: 10.1097/ACM.0b013e3181edfe13] [Medline: 20703150]
6. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Work Group for Evaluation and Implementation of Simulators and Skills Training Programmes. Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 2005 Dec;19(12):1523-1532. [doi: 10.1007/s00464-005-0384-2] [Medline: 16252077]
7. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012 Oct;99(10):1322-1330. [doi: 10.1002/bjs.8819] [Medline: 22961509]
8. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc 2003 Oct;17(10):1525-1529. [doi: 10.1007/s00464-003-0035-4] [Medline: 14502403]
9. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg 2003 Sep;73(9):712-716. [Medline: 12956787]
10. 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]
11. 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]

12. Chodos D, Stroulia E, King S. MeRiTS: simulation-based training for healthcare professionals. Stud Health Technol Inform 2011;163:125-131. [Medline: 21335774]
13. Cohen D, Sevdalis N, Patel V, Taylor M, Lee H, Vokes M, et al. Tactical and operational response to major incidents: feasibility and reliability of skills assessment using novel virtual environments. Resuscitation 2013 Jul;84(7):992-998. [doi: 10.1016/j.resuscitation.2012.12.011] [Medline: 23357703]
14. Dev P, Heinrichs WL, Youngblood P. CliniSpace: a multiperson 3D online immersive training environment accessible through a browser. Stud Health Technol Inform 2011;163:173-179. [Medline: 21335784]
15. Creutzfeldt J, Hedman L, Heinrichs L, Youngblood P, Felländer-Tsai L. Cardiopulmonary resuscitation training in high school using avatars in virtual worlds: an international feasibility study. J Med Internet Res 2013;15(1):e9 [FREE Full text] [doi: 10.2196/jmir.1715] [Medline: 23318253]
16. 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]
17. Funke K, Bonrath E, Mardin WA, Becker JC, Haier J, Senninger N, et al. Blended learning in surgery using the Inmedea Simulator. Langenbecks Arch Surg 2013 Feb;398(2):335-340. [doi: 10.1007/s00423-012-0987-8] [Medline: 22878596]
18. 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]
19. Caudell TP, Summers KL, Holten J, Hakamata T, Mowafi M, Jacobs J, et al. Virtual patient simulator for distributed collaborative medical education. Anat Rec B New Anat 2003 Jan;270(1):23-29 [FREE Full text] [doi: 10.1002/ar.b.10007] [Medline: 12526063]
20. Youngblood P, Harter PM, Srivastava S, Moffett S, Heinrichs WL, Dev P. Design, development, and evaluation of an online virtual emergency department for training trauma teams. Simul Healthc 2008;3(3):146-153. [doi: 10.1097/SIH.0b013e31817bedf7] [Medline: 19088658]
21. Knight JF, Carley S, Tregunna B, Jarvis S, Smithies R, de Freitas S, et al. Serious gaming technology in major incident triage training: a pragmatic controlled trial. Resuscitation 2010 Sep;81(9):1175-1179. [doi: 10.1016/j.resuscitation.2010.03.042] [Medline: 20732609]
22. Kurenov SN, Cance WW, Noel B, Mozingo DW. Game-based mass casualty burn training. Stud Health Technol Inform 2009;142:142-144. [Medline: 19377134]
23. Gutiérrez F, Pierce J, Vergara VM, Coulter R, Saland L, Caudell TP, et al. The effect of degree of immersion upon learning performance in virtual reality simulations for medical education. Stud Health Technol Inform 2007;125:155-160. [Medline: 17377256]
24. Alverson DC, Saiki SM, Kalishman S, Lindberg M, Mennin S, Mines J, et al. Medical students learn over distance using virtual reality simulation. Simul Healthc 2008;3(1):10-15. [doi: 10.1097/SIH.0b013e31815f0d51] [Medline: 19088637]
25. Wünschel M, Leichtle U, Wülker N, Kluba T. Using a web-based orthopaedic clinic in the curricular teaching of a German university hospital: analysis of learning effect, student usage and reception. Int J Med Inform 2010 Oct;79(10):716-721. [doi: 10.1016/j.ijmedinf.2010.07.007] [Medline: 20732829]
26. Horstmann M, Renninger M, Hennenlotter J, Horstmann CC, Stenzl A. Blended E-learning in a Web-based virtual hospital: a useful tool for undergraduate education in urology. Educ Health (Abingdon) 2009 Aug;22(2):269. [Medline: 20029750]
27. Lee CH, Liu A, Del Castillo S, Bowyer M, Alverson D, Muniz G, et al. Towards an immersive virtual environment for medical team training. Stud Health Technol Inform 2007;125:274-279. [Medline: 17377284]
28. Papakonstantinou D, Poulymenopoulou M, Malamateniou F, Vassilacopoulos G. Enabling the use of enhanced medical SOPs by an mLearning training solution. Stud Health Technol Inform 2013;190:86-88. [Medline: 23823384]
29. Kuhlen TW, Hentschel B. Quo vadis CAVE: does immersive visualization still matter? IEEE Comput Graph Appl 2014;34(5):14-21. [doi: 10.1109/MCG.2014.97] [Medline: 25248196]
30. Botella C, Rey A, Perpiñá C, Baños R, Alcañiz M, Garcia-Palacios A, et al. Differences on presence and reality judgment using a high impact workstation and a PC workstation. Cyberpsychol Behav 1999;2(1):49-52. [doi: 10.1089/cpb.1999.2.49] [Medline: 19178263]
31. Huerta R. Dissertations & Theses, The University of Texas - Pan American. Ann Arbor, MI: UMI Dissertations Publishing; 2012 Aug 01. Measuring the impact of narrative on player's presence and immersion in a first person game environment [MS disseration] URL: http://search.proquest.com/docview/1152549474 [accessed 2015-03-25] [WebCite Cache ID 6XIJRkLfn]
32. Welke TM, LeBlanc VR, Savoldelli GL, Joo HS, Chandra DB, Crabtree NA, et al. Personalized oral debriefing versus standardized multimedia instruction after patient crisis simulation. Anesth Analg 2009 Jul;109(1):183-189. [doi: 10.1213/ane.0b013e3181a324ab] [Medline: 19535709]
33. Catchpole K, Mishra A, Handa A, McCulloch P. Teamwork and error in the operating room: analysis of skills and roles. Ann Surg 2008 Apr;247(4):699-706. [doi: 10.1097/SLA.0b013e3181642ec8] [Medline: 18362635]

Abbreviations

3D: three-dimensional
IPS: immersive patient simulator
SL: Second Life
VW: virtual world

Edited by G Eysenbach; submitted 27.04.14; peer-reviewed by M Troconis, S Edelbring, SY Liaw; comments to author 13.12.14; revised version received 17.01.15; accepted 03.02.15; published 08.04.15.

Please cite as:
Kleinert R, Wahba R, Chang DH, Plum P, Hölscher AH, Stippel DL
3D Immersive Patient Simulators and Their Impact on Learning Success: A Thematic Review
J Med Internet Res 2015;17(4):e91
URL: http://www.jmir.org/2015/4/e91/
doi:10.2196/jmir.3492
PMID:25858862

Original Paper

Review and Evaluation of Online Tobacco Dependence Treatment Training Programs for Health Care Practitioners

Peter Selby[1], MBBS, CFPC, FCFP, DipABAM; Karina Goncharenko[1], MSc; Megan Barker[1], MA; Myra Fahim[1], MSc, BEd; Valerie Timothy[1], MDE; Rosa Dragonetti[1], MSc; Katherine Kemper[2], MBA; Marilyn Herie[3], PhD, RSW; J Taylor Hays[4], MD

[1]Centre for Addiction and Mental Health, Addictions Division, Toronto, ON, Canada
[2]Mayo Clinic, Global Bridges, Rochester, MN, United States
[3]Centennial College, Community Services, Toronto, ON, Canada
[4]Mayo Clinic, Nicotine Dependence Center, Rochester, MN, United States

Corresponding Author:
Peter Selby, MBBS, CFPC, FCFP, DipABAM
Centre for Addiction and Mental Health
Addictions Division
175 College Street
3rd Floor
Toronto, ON, M5T1P7
Canada
Phone: 1 416 535 8501 ext 36859
Fax: 1 416 599 3802
Email: peter.selby@camh.ca

Abstract

Background: Training health care professionals is associated with increased capacity to deliver evidence-based smoking cessation interventions and increased quit rates among their patients. Online training programs hold promise to provide training but questions remain regarding the quality and usability of available programs.

Objective: The aim was to assess the quality of English-language online courses in tobacco dependence treatment using a validated instrument.

Methods: An environmental scan was conducted using the Google search engine to identify available online tobacco dependence treatment courses. The identified courses were then evaluated using the Peer Review Rubric for Online Learning, which was selected based on its ability to evaluate instructional design. It also has clear and concise criteria descriptions to ensure uniformity of evaluations by trained experts.

Results: A total of 39 courses were identified, of which 24 unique courses were assessed based on their accessibility and functionality during the period of evaluation. Overall, the course ratings indicated that 17 of 24 courses evaluated failed to meet minimal quality standards and none of the courses evaluated could be ranked as superior. However, many excelled in providing effective navigation, course rationale, and content. Many were weak in the use of instructional design elements, such as teaching effectiveness, learning strategies, instructor’s role, and assessment and evaluation. Evaluation results and suggestions for improvement were shared with course administrators.

Conclusions: Based on the courses evaluated in this review, course developers are encouraged to employ best practices in instructional design, such as cohesiveness of material, linearity of design, practice exercises, problem solving, and ongoing evaluation to improve existing courses and in the design of new online learning opportunities.

(J Med Internet Res 2015;17(4):e97) doi:10.2196/jmir.3284

KEYWORDS

distance education; tobacco use; health care; smoking cessation; tobacco dependence; program evaluation; continuing medical education

Introduction

Tobacco use remains a global preventable cause of disease, disability, and death. It is estimated to kill 6 million people worldwide annually, most in the developing world where the prevalence of smoking remains high [1]. The World Health Organization's Framework Convention on Tobacco Control (FCTC-Article 14) highlights the need for countries to provide widely accessible cessation services to address the high level of global tobacco use. However, in many countries throughout the world, tobacco dependence treatment services are not well developed and the main pathway of accessing treatment for tobacco dependence is through the clinical interactions between a doctor and patient. There is evidence that training health care professionals in smoking cessation interventions is associated with increased quit rates in smokers [2] and that individuals who smoke are more likely to make an attempt to quit when advised to do so by a health care practitioner [3,4]. Despite the global health and financial burden of disease from tobacco use, the desire of most patients to quit, and routine contact between health care practitioners and patients, few practitioners are appropriately trained or feel confident to effectively treat tobacco addiction in their patients [4]. To address the high rates of tobacco prevalence and increase cessation levels among individuals who smoke, appropriate and effective training for health care practitioners in tobacco dependence treatment is pertinent.

Successful cessation outcomes can be directly correlated to training health care practitioners in evidence-based tobacco dependence treatment. A Cochrane Review [2] of studies measuring the effectiveness of tobacco dependence treatment training on successful cessation outcomes found that in nearly all the studies reviewed, smoking cessation activities of health care practitioners (including psychosocial and pharmacological interventions) increased dramatically posttraining. Moreover, a study by Olano-Espinosa and colleagues [5] found that training primary care health care practitioners in tobacco dependence treatment had a statistically significant effect on sustained abstinence after 6 months (2.1% in trained group vs 0.3% in the comparison group). These studies suggest that training health care practitioners in tobacco dependence treatment has a direct impact on increased levels of cessation activities among practitioners and successful quit attempts among patients. Given that health care practitioners in any clinical setting are well-suited to engage their patients in tobacco dependence treatment, training reinforces the notion to go beyond simply inquiring about tobacco use status and to instead offer evidence-based treatment. Online tobacco intervention training courses, similar to other e-learning programs, provide many potential benefits such as reaching greater numbers of practitioners from different disciplines, reducing the time required for practitioners to dedicate to improving skills and knowledge compared to in-class courses, and providing a cost-effective option to learning (less time away from patients, courses can be started and paused at any time, no travel required, smaller fees compared to a classroom-based course). Further evidence suggests that online continuing education courses, through incorporation of interactive components, are at least as effective as traditional classroom-based courses [6,7]. To increase reach and educate health care providers in effective tobacco interventions, there has been an increase in Internet-based instruction as part of continuing professional development. Online learning or e-learning is characterized by the use of telecommunication technology to deliver information for education and training and has emerged as the exemplar of modern education [8]. Studies have shown that offering health care providers online tobacco dependence intervention training that builds on current knowledge and skills improved the participants' attitudes and increased self-efficacy in delivering tobacco dependence treatment [9]. Not only did participants report significantly higher positive attitudes and improved self-efficacy for delivering tobacco dependence treatment services posttraining, but they also demonstrated increased delivery of these interventions with clients [9].

A common criticism of online courses is that they have inadequate evaluation mechanisms to enhance the learner's ability to apply the knowledge gained and instead merely provide facts. This is especially true when training health care practitioners who see themselves as problem solvers and want practical information to address problems commonly encountered in their practices. Unfortunately, there is no accreditation standard or similar quality measure that allows potential end users to know whether a course has been designed using the accepted best practices for online course design and educational theory. A thorough curation of available online tobacco cessation courses ensures more rigorous standards for delivery of tobacco cessation knowledge and best practices to health care providers via an online format. Therefore, as a first step in assessing online course quality, we set out to review and rank all available English-language online courses in tobacco addiction treatment training using a validated instrument. Secondary aims were to provide feedback to course developers about recommendations for course improvement.

Methods

Online Tobacco Course Identification (Environmental Scan)

We conducted a basic search on the Google search engine using variations of "online smoking cessation courses for health care practitioners," "smoking cessation course," and "smoking cessation online course/training" keywords and followed links to courses and online resources for practitioners and clients related to smoking cessation embedded in government-funded or -supported websites (eg, Ontario Tobacco Research Unit course) or resource sections of other online course websites. The search to identify online courses available to practitioners seeking an e-learning module was conducted in August 2012. An online course description, estimated time to complete, developer information, the language of instruction, availability of continuing medical education (CME) credits, and course fees were included in the analysis (Multimedia Appendix 1).

Online Tobacco Course Evaluation

Once the courses were identified, the course director or contact person listed for the course was sent an email invitation to participate (Multimedia Appendix 2). We sent 2 emails as reminders. Evaluation of online-based training courses was then carried out by the evaluation team of the Peer Review Rubric for Online Learning (developed by Towson University Faculty in Maryland) [10]. This rubric was chosen based on its ability to evaluate both instructional design and instructional content of an online course, ability to apply the rubric to a variety of online learning programs, and its clear and concise criteria descriptions to ensure uniformity of evaluations. Moreover, the developers trained raters in the use of the instrument thereby reducing the risk of bias by the investigators and authors. The course evaluations addressed the extent to which each online course adhered to best practices in online teaching/learning, including instructional design enhancements such as the presence of case studies, practice exercises, clinical simulations and demonstrations, video demonstration, and group discussion or consultation, as well as content subject matter relating to the science, policy, and practice of tobacco prevention and smoking cessation. Each of the 16 categories evaluated by the rubric (Textbox 1) allowed for a minimal score of 0 (=not included) and a maximal score of 4 (=superior). The overall rating score allowed for a maximum of 64 points: 54-64=superior, 48-53=above average-good, 41-47=average-OK, 35-40=a start, but needs polish, ≤34=redo entirely. The developers of the rubric opted to use a nonstandardized measure to score the courses because the scale was based on the points within the rubric that were assigned to a percentage range. For example, superior points were based on achieving 85% of the maximum of 64 points (range 54-64 points). The reviewers also established a minimal changes score by highlighting categories where minimal changes could provide maximum points. Detailed examples and comments were then provided on possible changes to enhance the course. Developers of online courses who agreed to participate in this evaluation had the option to receive individualized feedback and information about additional resources/contacts if they wished to engage in quality improvement of their online modules. Those interested in discussing the results of the evaluation in more detail were invited to participate in an in-depth discussion via teleconference.

The external course evaluation/rating team had approximately 15 years of experience in the use of the rubric and established 95% reliability based on few differences in rating scores. The course evaluation/rating team was comprised of senior authors for the rubric who are nationally certified as Peer Reviewers and Master Reviewer. Furthermore, the rating team had no previous knowledge of the content of the courses being considered for review and no prior experience working for or with the authors of this paper.

Textbox 1. Categories of evaluation outlined by the Peer Review Rubric for Online Learning.

1. Navigation
2. Course rationale
3. Learning and teaching theories
4. Instructional design
5. Goals and objectives
6. Learning strategies
7. Content
8. Interactivity
9. Use of mediated resources and Web
10. Assessment and evaluation
11. Internal organization and consistency
12. Responsiveness to learner’s needs
13. Instructor’s role
14. Teaching effectiveness
15. How to get help
16. Esthetics

Results

We identified 39 courses in English that met our initial criteria. Each of the 39 online courses ranged in length from 1 hour to 20 hours to complete. A total of 27 courses were found to be unique, accessible, and functioning by the end of April 2013. Of these, 3 courses had partial issues of accessibility, such as virus interference and inability to continue to the next module or progress with course material. As a result, we completed evaluations for 24 unique courses (Table 1). Overall course ratings revealed that the majority of courses (17 courses) failed to meet minimal quality standards and needed complete revision, 2 courses were below average (“need polish”), 2 courses were “average-OK,” and only 3 courses ranked as “above average-good” with none being “superior” (Figure 1).

Table 1. Course scores as assessed with the Peer Review Rubric for Online Learning.

Course	Evaluation category (score range 0-4)																Course score (max=64)	Minimal changes score (max=64)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		
1	3	4	4	4	3	3	4	3	2	2	3	3	4	2	3	4	51	61
2	2	4	4	2	1	3	4	3	3	2	2	3	3	2	3	2	43	51
3	4	2	2	1	0	0	2	2	2	2	0	0	1	0	0	3	21	29
4	4	4	3	1	2	1	3	1	1	2	2	2	1	0	0	2	29	43
5	4	4	3	1	0	1	3	1	1	1	1	2	1	0	0	2	25	40
6	2	3	0	1	0	1	1	0	1	2	1	1	1	0	0	1	15	26
7	3	2	3	3	2	1	3	1	1	2	2	2	1	1	1	3	31	49
8	3	4	3	3	3	2	3	2	2	1	3	3	3	0	3	3	41	48
9	2	3	3	2	1	1	3	1	2	1	3	1	1	1	2	2	29	40
10	2	3	2	1	1	0	2	1	2	1	2	2	2	2	1	1	24	37
11	4	4	4	4	4	2	3	3	3	2	3	3	1	3	4	4	51	56
12	2	4	2	3	1	1	3	1	1	1	3	1	1	1	1	2	28	42
13	3	2	0	2	1	0	1	0	1	1	1	0	1	0	1	1	15	18
14	3	1	3	1	1	1	3	2	2	1	1	2	1	2	0	3	27	40
15	2	4	4	3	1	2	3	2	2	1	1	2	1	1	3	1	33	46
16	4	4	4	4	4	3	4	4	2	2	3	3	2	2	0	3	48	58
17	2	3	1	1	1	0	1	0	0	1	0	0	1	0	2	1	14	25
18	3	4	1	1	0	1	3	1	1	1	3	1	1	0	2	2	25	32
19	3	4	3	2	4	1	3	1	2	3	4	2	1	2	1	2	38	49
20	3	4	3	1	2	2	3	2	2	2	2	2	1	2	4	3	38	47
21	2	4	3	1	2	1	3	2	2	0	1	1	1	0	1	3	27	33
22	3	2	3	2	2	2	2	1	1	0	3	3	1	0	3	2	30	42
23	2	4	0	2	3	1	2	2	1	1	2	0	1	1	2	1	25	47
24	2	4	2	1	3	0	3	2	0	1	1	0	1	2	4	2	28	34

To facilitate improvement of the online courses without extensive redevelopment, moderate and quickly implementable amendments were suggested for each of the courses evaluated. These minimal changes were determined at the discretion of the rubric experts. An example of a minimal change suggested by the reviewers was to offer more for learner-led opportunities, such as discussion boards or self-reflective questions to maximize transformational learning. The 3 courses that were rated most highly required only minor changes to become “superior” (Figure 2). Furthermore, with minimal changes, most of the 17 courses with unacceptably low scores could improve their quality rating: 4 courses could be improved to “needs polish,” 6 courses could improve to “average-OK,” and 4 courses could achieve a rating of “above average-good.”

Figure 3 presents mean scores for each category assessed by the rubric. On average, most courses scored well in course rationale, navigation, and content, whereas teaching effectiveness, learning strategies, assessment and evaluation, and instructor’s role proved to be challenging and underdeveloped in most courses. Overall, on the scale of 0-4, most courses scored higher than 2 in 5 categories, scored exactly 2 in 2 categories, and received a score less than 2 on the scale in 9 categories.

Online training programs are often defined by 2 types of e-learning: asynchronous and synchronous. Asynchronous e-learning occurs when students begin and end a training program at different times according to their own schedule and personal preferences. Common features of asynchronous learning include online bulletin and discussion boards, group forums, and self-directed learning. Approximately 92% (22/24) of courses reviewed in this evaluation were categorized as asynchronous. Synchronous e-learning occurs when remote students enroll in a course that follows a specific schedule as outlined by the course instructor. Failure to virtually attend the course or complete assignments according to the scheduled deadlines could result in negative course outcomes. Common features of synchronous learning include live facilitated discussion via whiteboards, virtual classrooms that promote instructor-to-learner and learner-to-learner engagement, and scheduled learner evaluation. Approximately 8% (2/24) of the courses reviewed were categorized as synchronous. Given the number of asynchronous courses reviewed, it is no surprise that 96% (23/24) of the courses evaluated were categorized as self-study. Only 1 course could be categorized as offering collaborative learning because it encouraged learner-to-learner participation in a synchronous virtual classroom in which students were asked to respond to one another's questions and provide feedback in real time.

Of the 24 courses included in the review, 17 agreed to be evaluated and were provided with individualized feedback and information on additional resources that could improve their course overall. When providing feedback, all courses participating in the review process were offered the opportunity to engage in a more in-depth discussion regarding their evaluation by teleconference. Only 1 of the 17 courses chose the option to further discuss their course review.

Figure 1. Distribution of courses evaluated by course quality without suggested changes.

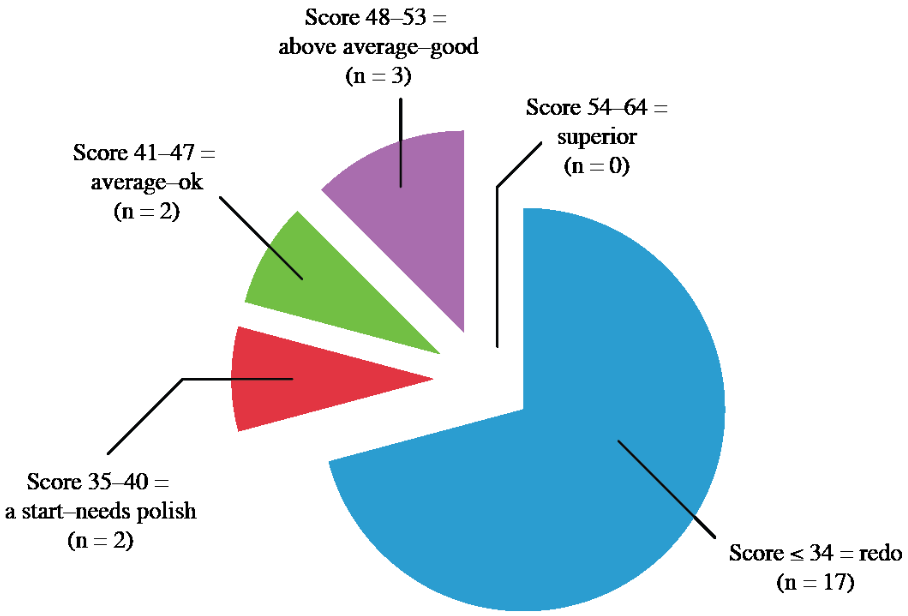


Figure 2. Distribution of courses evaluated by course quality with suggested changes.

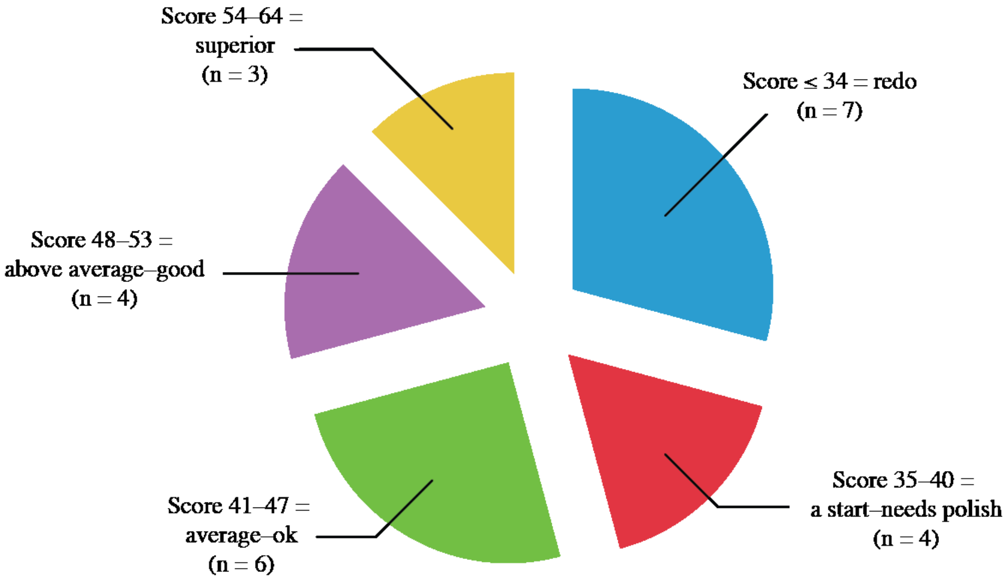
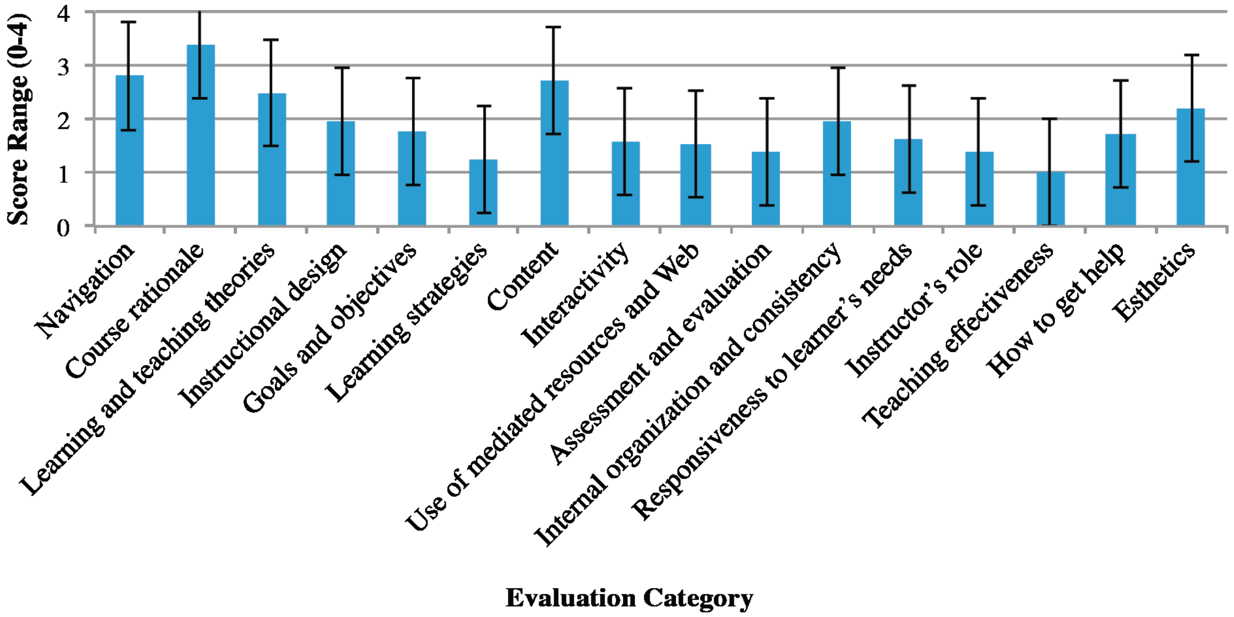


Figure 3. Mean category scores for 24 courses based on the 16 categories evaluated. Error bars indicate standard deviation.



Discussion

Tobacco use and dependence remains the most important cause of preventable death and morbidity worldwide. Despite this fact, many health professionals do not provide effective treatment to tobacco users, in part because of inadequate knowledge and skills in treating the problem. Online education providing instruction in tobacco dependence treatment is one important way to build the necessary capacity to provide evidence-based treatment for tobacco dependence. The online learning environment overcomes many of the barriers found in traditional instruction by expanding the reach and convenience of the instructional material for adult learners. However, adult learners also need to be able to readily access high-quality online learning opportunities to maximize the limited time available for acquiring new knowledge and skills. Unfortunately, the main finding of our review of available online courses for tobacco dependence treatment was not encouraging in this regard. Out of the initial 39 courses identified, a mere 24 courses were able to be accessed for evaluation. The primary obstacle to these courses was inability to access course content due mostly to inadequate course maintenance (ie, ensuring the webpage is virus-free, links are in working condition). Among the 24 courses we evaluated, the majority (17/24) require complete revision to meet minimal quality standards for best practices and optimal instructional design. Fortunately, we also found that there would be a significant positive impact on course quality (an increase of 25% in quality rating scores) with modest revisions. This suggests that there exists a relatively easy way to improve most available courses to meet the average industry standard. A detailed analysis of each of the 16 categories of the assessment rubric makes it clear that most online courses will obtain the greatest benefit by improving in the areas of teaching effectiveness, learning strategies, instructor's role, and assessment and evaluation. Course developers are encouraged to focus on a wider range of design elements necessary for a successful course (eg, interactivity, aligning learning strategies).

A review of e-learning literature for health care practitioners suggests areas for quality improvement comparable to those found in our study [11]. For example, the review noted that learner testing and assessment is crucial to ensure that information has been absorbed and will be applied, and learners need to receive feedback on their performance. Among the lowest ranking areas found in our study was “assessment and evaluation.” Similarly, low-ranking categories found in our study were instructor's role, teaching effectiveness, and responsiveness to learner's needs. The review also found that instructors and developers need to be flexible to adapt to the needs of learners and that learner participation and interaction needs to be supported and encouraged similar to that of traditional classroom environments [11]. Collaboration between content, pedagogy, and technology is needed to keep learners engaged [11]. We found that course interactivity, learning strategies, and instructional design—features that enhance learner engagement—were lacking for most tobacco dependence online training.

As a result of our findings of overall poor quality of online courses, we suggest that course developers raise the quality and educational value of their courses by using the following strategies. First, subject matter experts must engage experts in instructional design to ensure course content is palatable for an online format and aligned with learning objectives. Second, interactive learning exercises need to be incorporated to maintain learner engagement and retention of course content [12]. Third, it is important to assess learners in a way that can facilitate knowledge transfer to real-world scenarios [12]. Simply providing learners with knowledge-based multiple-choice questions can be useful in particular educational settings; however, in CME courses it is crucial that application of knowledge is assessed. Finally, course designers need to remember that online courses require constant oversight and continuous improvement to keep course content and interactive learning activities current and fresh.

The potential to engage in meaningful reflection and optimize the multimedia potential to enhance cognitive, affective, and kinesthetic learning is the promise of e-learning. When developing an online course, it is important to remember that simply uploading text, charts, visuals, or a PowerPoint presentation to an online environment does not create an e-learning module. Unfortunately, this was the primary design approach for many of the courses reviewed. Modern education is characterized by the employment of multiple modalities to engage a wide variety of learners. Thus, effort needs to be made by course developers to utilize different learning approaches beyond didactic text, such as case simulations, quizzes, interactive learning objects, and discussion forums. Moreover, instructional design principles suggest that there needs to be alignment of the learning objectives, the content (including subcontent), the case scenarios, and assessments. Often, we observed misalignment of these essential elements—a design flaw risking disengagement of the learner from the material irrespective of the quality of the content. Additionally, pilot testing the e-learning module to evaluate both content and instructional design prior to implementation and widespread dissemination is imperative. Pilot testing the module in a sample of the target audience is an opportunity to garner valuable feedback and allow course developers to make improvements and enhance the module for future learners.

There are several limitations to our study. Although the rubric we chose was deemed to be the most thorough in assessing best practices and instructional design out of the 10 rubrics reviewed, there are limitations in its use. The rubric was designed for use in higher education online courses based on the assumption of a course management system that allows for learner interaction, such as learner-to-learner, learner-to-content, and learner-to-instructor interaction. Unfortunately, almost none of the courses we evaluated fostered such interaction. Although the rubric was designed to capture both design and content, there seemed to be a stronger focus on the e-learning environment and less on the content. This may prove to be a limitation as diverse delivery needs may be required for different content approaches. Both content and delivery are important to the outcomes, but the rubric assumes a higher weighting on delivery than content. In addition, replicating this evaluation may be challenging because the rubric requires trained, expert reviewers who have had years of experience working in tandem and cannot be generally applied for use by untrained reviewers. This also impacts the minimal changes scoring because the expert reviewer team relies on 15 years of experience in online course design and delivery to determine the minimal changes score. This score was determined by highlighting categories where minimal changes would provide maximum points. These subjective observations may prove to be problematic when attempting to quantify an overall numeric score. We also excluded from our review almost 40% (15/39) of the online courses we identified. Because these courses could not be accessed for meaningful review, we cannot comment on their quality or whether their inclusion in our study would have resulted in different conclusions. Finally, we cannot be certain that we identified all available English-language online learning courses because there is no comprehensive source for all courses on tobacco dependence treatment instruction. However, our search strategy mirrors that of most experienced online learners when seeking instructional content.

Our main conclusion is that there are important quality gaps between available online courses for tobacco dependence treatment instruction and the high quality that all courses should strive to achieve. Our evaluation makes it clear that there is a widespread lack of well-designed online continuing education courses in tobacco dependence treatment based on an analysis of instructional design quality. However, optimizing currently available online learning tools may not require extensive redesign or costly effort. Implementing modest changes will improve the quality of most existing courses to at least an average quality level and courses with average or above average quality may achieve superior-level quality with similar modest revision. The design elements that the majority of the courses lack are cohesiveness between the different module components and the linearity of the design. Because most courses in tobacco dependence treatment will be used by adult learners who are already working in a health care profession, course designers must also provide information succinctly and provide opportunity within the course for practice and problem solving. Simply providing abundant erudite content will not meet the needs of most adult learners. Having identified quality gaps in current online learning in tobacco dependence treatment instruction as well as feedback for developers of existing online courses, we would encourage course developers to employ these best practices and feedback for improvements to existing courses and in the design of new online learning opportunities. We believe there is a great unmet need for quality online education and that using instructional design principles could ensure greater impact of any content made available to online learners.

Acknowledgments

Global Bridges Mayo Clinic is funded by a grant from the Pfizer Medical Group. K Goncharenko and V Timothy have since moved on from the Centre for Addiction and Mental Health to work in the Private Sector.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Identified online tobacco cessation training programs.

[PDF File (Adobe PDF File), 133KB - jmir_v17i4e97_app1.pdf]

Multimedia Appendix 2

(Sample) Request for participation in the online course evaluation.

[PDF File (Adobe PDF File), 68KB - jmir_v17i4e97_app2.pdf]

References

1. Shafey O, Ericksen M, Ross H, MacKay J. The Tobacco Atlas. 3rd ed. Atlanta, GA: American Cancer Society; 2009.
2. Carson KV, Verbiest MEA, Crone MR, Brinn MP, Esterman AJ, Assendelft WJJ, et al. Training health professionals in smoking cessation. Cochrane Database Syst Rev 2012;5:CD000214. [doi: 10.1002/14651858.CD000214.pub2] [Medline: 22592671]
3. Thorndike AN, Regan S, Rigotti NA. The treatment of smoking by US physicians during ambulatory visits: 1994 2003. Am J Public Health 2007 Oct;97(10):1878-1883. [doi: 10.2105/AJPH.2006.092577] [Medline: 17761570]
4. Maynard A, Metcalf M, Hance L. Enhancing readiness of health profession students to address tobacco cessation with patients through online training. Int J Med Educ 2012 Mar 26;3:57-62. [doi: 10.5116/ijme.4f70.bd8b]
5. Olano-Espinosa E, Matilla-Pardo B, Minué C, Antón E, Gómez-Gascón T, Ayesta FJ. Effectiveness of a health professional training program for treatment of tobacco addiction. Nicotine Tob Res 2013 Oct;15(10):1682-1689. [doi: 10.1093/ntr/ntt040] [Medline: 23547275]
6. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
7. Lam-Antoniades M, Ratnapalan S, Tait G. Electronic continuing education in the health professions: an update on evidence from RCTs. J Contin Educ Health Prof 2009;29(1):44-51. [doi: 10.1002/chp.20005] [Medline: 19288566]
8. Sun P, Tsai RJ, Finger G, Chen Y, Yeh D. What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. Computers & Education 2008 May;50(4):1183-1202. [doi: 10.1016/j.compedu.2006.11.007]
9. Carpenter KM, Carlini BH, Painter I, Mikko AT, Stoner SA. Refer2Quit: impact of Web-based skills training on tobacco interventions and quitline referrals. J Contin Educ Health Prof 2012;32(3):187-195. [doi: 10.1002/chp.21144] [Medline: 23008081]
10. Ashcraft M, McMahon J, Lesh S, Tabrizi M. Towson University. 2003. Peer review for online learning URL: http://pages.towson.edu/mcmahon/peerreview/On-linerubric.pdf [accessed 2013-10-16] [WebCite Cache ID 6KPmiIXbW]
11. Childs S, Blenkinsopp E, Hall A, Walton G. Effective e-learning for health professionals and students--barriers and their solutions. A systematic review of the literature--findings from the HeXL project. Health Info Libr J 2005 Dec;22 Suppl 2:20-32. [doi: 10.1111/j.1470-3327.2005.00614.x] [Medline: 16279973]
12. Chan CH, Robbins LI. E-Learning systems: promises and pitfalls. Acad Psychiatry 2006;30(6):491-497. [doi: 10.1176/appi.ap.30.6.491] [Medline: 17139020]

Abbreviations

CME: Continuing Medical Education

Edited by G Eysenbach; submitted 01.02.14; peer-reviewed by H Travis, A Kirkwood; comments to author 30.07.14; revised version received 01.10.14; accepted 20.12.14; published 17.04.15.

Please cite as:
Selby P, Goncharenko K, Barker M, Fahim M, Timothy V, Dragonetti R, Kemper K, Herie M, Hays JT
Review and Evaluation of Online Tobacco Dependence Treatment Training Programs for Health Care Practitioners
J Med Internet Res 2015;17(4):e97
URL: http://www.jmir.org/2015/4/e97/
doi:10.2196/jmir.3284
PMID:25887187

Review

Globalization of Continuing Professional Development by Journal Clubs via Microblogging: A Systematic Review

Matthew John Roberts[1,2,3,4], BSc, MB BS; Marlon Perera[1,3,4], BmedSci, MB BS; Nathan Lawrentschuk[5,6,7], MB BS, FRACS, PhD; Diana Romanic[5,6,7], BA(psyc), BExSc; Nathan Papa[5], MB BS, MEpi; Damien Bolton[5,6], MB BS, MD, FRACS

[1]School of Medicine, The University of Queensland, Brisbane, Australia
[2]Centre for Clinical Research, The University of Queensland, Brisbane, Australia
[3]Mackay Base Hospital, Mackay, Australia
[4]Urological Society of Australia and New Zealand, Edgecliff, Australia
[5]University of Melbourne, Department of Surgery, Austin Hospital, Melbourne, Australia
[6]Olivia Newton-John Cancer Research Institute, Austin Health, Melbourne, Australia
[7]Peter MacCallum Cancer Centre, Division of Cancer Surgery, Melbourne, Australia

Corresponding Author:
Matthew John Roberts, BSc, MB BS
School of Medicine
The University of Queensland
RBWH Campus
Herston
Brisbane, 4029
Australia
Phone: 61 0422378975
Fax: 61 7 3346 5509
Email: m.roberts2@uq.edu.au

Related Articles:

Letter: http://www.jmir.org/2015/9/e216/

Letter: http://www.jmir.org/2015/9/e217/

Abstract

Background: Journal clubs are an essential tool in promoting clinical evidence-based medical education to all medical and allied health professionals. Twitter represents a public, microblogging forum that can facilitate traditional journal club requirements, while also reaching a global audience, and participation for discussion with study authors and colleagues.

Objective: The aim of the current study was to evaluate the current state of social media–facilitated journal clubs, specifically Twitter, as an example of continuing professional development.

Methods: A systematic review of literature databases (Medline, Embase, CINAHL, Web of Science, ERIC via ProQuest) was performed according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A systematic search of Twitter, the followers of identified journal clubs, and Symplur was also performed. Demographic and monthly tweet data were extracted from Twitter and Symplur. All manuscripts related to Twitter-based journal clubs were included. Statistical analyses were performed in MS Excel and STATA.

Results: From a total of 469 citations, 11 manuscripts were included and referred to five Twitter-based journal clubs (#ALiEMJC, #BlueJC, #ebnjc, #urojc, #meded). A Twitter-based journal club search yielded 34 potential hashtags/accounts, of which 24 were included in the final analysis. The median duration of activity was 11.75 (interquartile range [IQR] 19.9, SD 10.9) months, with 7 now inactive. The median number of followers and participants was 374 (IQR 574) and 157 (IQR 272), respectively. An overall increasing establishment of active Twitter-based journal clubs was observed, resulting in an exponential increase in total cumulative tweets (R^2=.98), and tweets per month (R^2=.72). Cumulative tweets for specific journal clubs increased linearly, with @ADC_JC,

@EBNursingBMJ, @igsjc, @iurojc, and @NephJC, and showing greatest rate of change, as well as total impressions per month since establishment. An average of two tweets per month was estimated for the majority of participants, while the “Top 10” tweeters for @iurojc showed a significantly lower contribution to overall tweets for each month ($P<.005$). A linearly increasing impression:tweet ratio was observed for the top five journal clubs.

Conclusions: Twitter-based journal clubs are free, time-efficient, and publicly accessible means to facilitate international discussions regarding clinically important evidence-based research.

(J Med Internet Res 2015;17(4):e103) doi:10.2196/jmir.4194

KEYWORDS

journal club; social media; continuing medical education; continuing professional development; systematic review

Introduction

Journal clubs are a well-established method to facilitate interactive peer review and critical thinking in clinical evidence-based medical education [1,2]. Furthermore, journal clubs provide a forum for academic debate and professional networking. Skills learned in critical analysis and literature appraisal skills are crucial in continuing professional development, in order to exercise best practices by any medical student or senior attending/consultant. Components of an effective journal club include regular and appropriate timing, high attendance rate (compulsory or incentive-based), nominated chairman, literature aligned with the journal club goals and reviewed prior to journal club session, and continuing professional development [3-6]. Unfortunately, many of these prerequisites act as limiting factors in a busy clinical practice.

An unprecedented expansion in the medical use of social media, such as Twitter, Facebook, LinkedIn, and YouTube, has followed the uptake seen with the general public. Twitter is a public, microblogging forum where users (each with a unique handle, eg @username) upload short messages comprising a maximum of 140 characters, with/or attached photos, documents, and links to other media such as videos, presentations, or journal articles. In most cases, these “tweets” are linked to a theme, often centralized using a “#” (hashtag, eg #twitter) for easy view and discovery by other users. Users are also able to “follow” the tweets of other users. Specific events (eg, medical conferences, public sporting events) often promote an official hashtag to allow users to follow all discussion relating to the event [7]. Dedicated explanation of social media apps in medicine and health care is available, with many falling under the broad banner of Free Open Access Medical education (FOAM) [8,9].

When combined with Twitter, journal clubs are able to function in a similar way to traditional journal clubs, with the advantage of a global audience and participation for discussion. Twitter-based journal clubs are able to be easily linked using a hashtag (eg, #...jc), allowing anyone to follow and contribute with a unique identifiable username. A central moderator is able to inform followers of the article to be discussed well ahead of time for perusal. Furthermore, authors of discussed articles are often invited as participants, enabling real-time interaction. Online journal clubs allow for international and increased participation, even when used with other less mainstream platforms [10,11]. Additional benefits of Web-based journal clubs include immediate feedback and discussion with authors and colleagues, as well as enhanced publication dissemination [1,12].

The aim of the current study was to evaluate the current state of journal clubs facilitated by social media, specifically Twitter, as an example of continuing professional development and through a systematic review process.

Methods

Overview

A systematic review was undertaken based on guidelines outlined by the Cochrane Collaboration and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [13].

Published Literature Search

A systematic literature search was performed using literature databases (Medline via OVID, Embase, CINAHL, Web of Science, ERIC via ProQuest) using synonyms relating to “social media” and “journal club” in November 2014 (see Multimedia Appendix 1) [14,15]. Reference lists of the identified articles were also searched. Initially, synonyms relating to medical education were included but were later removed due to overexclusion of potential manuscripts. Included articles were published reports on the use of social media (specifically Twitter in order to investigate a specific platform) as a means of facilitating journal clubs. Following exclusion of duplicates, irrelevant articles, and abstracts of conference proceedings based on citations and full text, remaining articles meeting inclusion criteria were reviewed for methodology and summarized. Article selection was performed by 2 independent evaluators (MR, MP) and any discrepancies resolved.

Twitter Hashtag Search

Following review of published reports, Twitter was by far the most popular and commonly used social media outlet for journal clubs. A systematic search of Twitter was performed to identify all relevant hashtags to be included in the current study, including using the search box and the terms “journal club” and “jc”, as well as reviewing the users who were following all identified journal clubs, initially using those identified in the literature search. Journal club selection was performed by 2 independent evaluators (MR, MP), and any discrepancies were resolved.

Inclusion criteria for hashtag analysis included journal clubs related to health care. Hashtags were excluded from analysis if they were not Twitter-based journal clubs, English-speaking, if the hashtag was not used completely for the purpose of a journal club, or if the hashtag represented institutional or private journal clubs. A final search was performed on the Symplur website, which tracks social media related to health care.

Data Extraction

Basic Twitter-based data was extracted from the relevant journal club Twitter-based websites. If available, such information included speciality, location, journal club tweets, and Twitter followers. Journal club tweets refer to the number of posts generated by a single journal club account. Data extraction from Symplur was achieved by searching each relevant hashtag (#). Data extracted from Symplur included hashtag commencement date, hashtag inactivation date, total tweet count, tweet count per month, number of tweeters, and total impressions. “Hashtag activation date” was defined as first month with greater than 5 tweets with the relevant hashtag. “Hashtag inactivation date” was defined as the starting point of 3 consecutive months with <6 tweets per month. Tweet count refers to the total number of posts containing the relevant hashtag. Number of tweeters was defined as the number of unique individual Twitter accounts that generated a post containing the relevant hashtag. Impressions, or reach, refer to the number of Twitter users using a particular hashtag and the sum of their respective followers—thus a surrogate for the number of users exposed to a particular hashtag. Detailed statistics of the top five journal clubs (@NephJC, @igsjc, @EBNursingBMJ, @iurojc, and @ADC_JC), as determined by rate of increase in cumulative tweets, were retrieved for each month from Symplur.

Statistical Analysis

Data from published manuscripts were insufficient for compilation, so a descriptive analysis was performed. Twitter hashtag data were collated and analyzed using a Microsoft Excel 2003 database. Figures were created using STATA v.12.0 SE.

Results

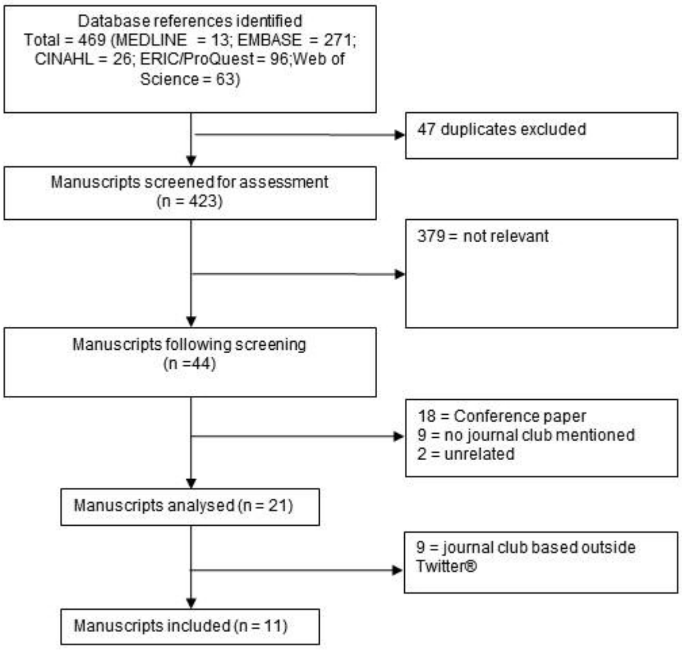
Literature Search

The final search strategy resulted in retrieval of 469 citations, including 47 duplicates, 381 irrelevant citations, 18 conference abstracts, 9 citations unrelated to journal clubs, and 9 based on blogs or other resources (Figure 1). Eleven manuscripts were included and analyzed [1,11,16-24], which included reference to five Twitter-based journal clubs (#ALiEMJC, #BlueJC, #ebnjc, #urojc, #meded).

Three manuscripts referred to #ALiEMJC/@AnnalsofEM [16-18] (which reported the social media responses to journal club discussions), four to #BlueJC/@bluejchost [1,19-21] (comprising a narrative description [1] with accompanying letter to the editor and author reply [19,20], and inclusion in a narrative review [21]), one to #ebnjc/@EBNursingBMJ in a regional nursing publication raising awareness for #ebnjc [22], two to #urojc/@iurojc [11,23] (comprising a narrative description and analysis of the first 12 months [11] and letter to the editor regarding a recently discussed manuscript [23]), and one to #meded/@JournalGIM (as an editorial providing a narrative description [24]).

Of the manuscripts reviewed, three journal clubs were officially affiliated with peer-reviewed journals (#ALiEMJC/@AnnalsofEM to *Annals of Emergency Medicine*, #BlueJC/@bluejchost to *BJOG An International Journal of Obstetrics and Gynaecology*, #meded/@JournalGIM to *Journal of General Internal Medicine*), while the remainder have grown out of specialty interest groups and as yet appear to have no official alignment.

Figure 1. Flowchart of study selection as per the PRISMA statement.



Twitter Hashtag Search

Following the hashtag and Twitter-based journal club search, 34 potential hashtags/accounts were collated. Following review, 10 hashtags and associated journal clubs were excluded: three due to multiple uses for particular hashtag, two hashtags were related to private or institutional journal clubs, four were excluded due to complete inactivity or commenced within 1 month of assessment, and one was excluded due to non-English language. Of the 24 included hashtag accounts, the median months active was 11.75 (interquartile range [IQR] 19.9, SD 10.9; Table 1). Symplur analytic data were unavailable for three hashtag-related journal clubs.

Table 1. Demographic information for Twitter-based journal clubs, including specialty, frequency, commencement, inactive status, and associated peer-reviewed publications.

Hashtag, citation	Associated Twitter user	Speciality	Frequency	Commencement	Inactive date[a]
#1carejc[b]	@amcunningham	Primary care	NR	15/04/2013	5/6/2014
#ACCJournalclub[b]	@ACCinTouch	Cardiology	NR	29/3/2014	29/3/14
#ADC_JC	@ADC_JC	Pediatrics	NR	28/5/2014	Current
#ALiEMJC [16-18]	@AnnalsofEM	Emergency medicine	NR	15/11/2013	Current
#ambjc[b]	@ambjournalclub	Emergency medicine	Bimonthly	1/7/2011	11/11/13
#bluejc[b] [1]	@bluejchost	Women's health	NR	30/3/2013	Current
#cpjc[b]	@clinpsyJC	Psychology	NR	1/7/11	1/10/13
#ebnjc [22]	@EBNursingBMJ	Nursing	Bimonthly	5/1/2010	Current
#GeriMedJC	@gerimedJC	Geriatric medicine	Monthly	1/7/2014	Current
#HEJC	twubs/HEJC	Health economics	NR	1/9/2012	Current
#hpmJC	@hpmjc	Hospice, palliative medicine/care	NR	1/03/14	Current
#IGSJC	@igsjc	General surgery	Monthly	5/2/2014	Current
#jamapedsjc	@jamapeds	NR	21/10/2014	8/8/14	Current
#JC_StE[b]	@JC_StE	Emergency medicine	Bimonthly	1/10/2012	1/10/13
#MedEd[c] [24]	@JournalGIM	General medicine	NR	NR	
#microtwjc	@microtwjc	Microbiology	NR	11/9/2013	Current
#NephJC	@Nephjc	Nephrology	NR	20/4/2014	Current
#PGHANJC	@BSPGHAN	Pediatric gastroenterology	NR	2/6/2014	
#PHTwitJC	@PHTwitJC	Public health	Monthly	1/7/2011	1/8/13
#psychjc	@psychiatryjc	Psychiatry	Monthly	28/9/2014	Current
#rsjc	@respandsleepjc	Respiratory	Monthly	26/6/2014	Current
#swjcchat[b]	@swjcchat	Social work	Bimonthly	7/7/2013	1/12/13
#twitjc[b]	@twitjournalclub	General medicine	Bimonthly	1/5/2011	1/12/13
#urojc [11]	@iurojc	Urology	Monthly	1/10/2012	Current

[a] If applicable, defined as 3 consecutive months of fewer than 5 tweets per month.

[b] Inactive journal clubs.

[c] #MedEd is not a unique hashtag for this journal club—it is also used for discussion among other medical educators.

Online Journal Club Activity

Of the 24 included journal club–related hashtags, the median number of followers was 374 (IQR 574) with a median number of active participants of 157 (IQR 272). Monthly activity was calculated with a median "tweets per month" of 203 (IQR 317) and median "impressions per month" of 165,538 (IQR 504,654). Following inception, and as of October 30, 2014, seven of the included journal clubs had become inactive (#1carejc, #ambjc, #cpjc, #JC_StE, #PHTwitJC, #swjcchat, #twitjc).

Overall, after exclusion of inactive journal clubs, an increasing establishment of Twitter-based journal clubs (Figure 2) was observed. Furthermore, an exponential increase in cumulative tweets was observed (R^2=.98) as well as tweets per month (R^2=.72).

For specific journal clubs, a continual increase in cumulative tweets in the early (<24 months) stages was observed. Specifically, the linearly modeled (all R^2>.95) increase in tweets was estimated as being greatest for @NephJC (722 tweets/month), @igsjc (613 tweets/month), @EBNursingBMJ (417 tweets/month), @iurojc (345 tweets/month), and @ADC_JC (255 tweets/month). These trends were also observed in the impressions/month rankings (Table 2). The two journal clubs that have previously been the longest running but now inactive, @twitjournalclub and @PHTwitJC, showed an inverse exponential cumulative tweet relationship following very high initial activity. Individually, no definite trends in tweets per month were observed, owing to large fluctuations in monthly activity.

Table 2. Twitter-based journal club performance, incorporating standard metrics such as tweets and followers, as well as overall tweets, participants and impressions, with calculated monthly mean tweets and impressions relating to each journal club as defined by Symplur.

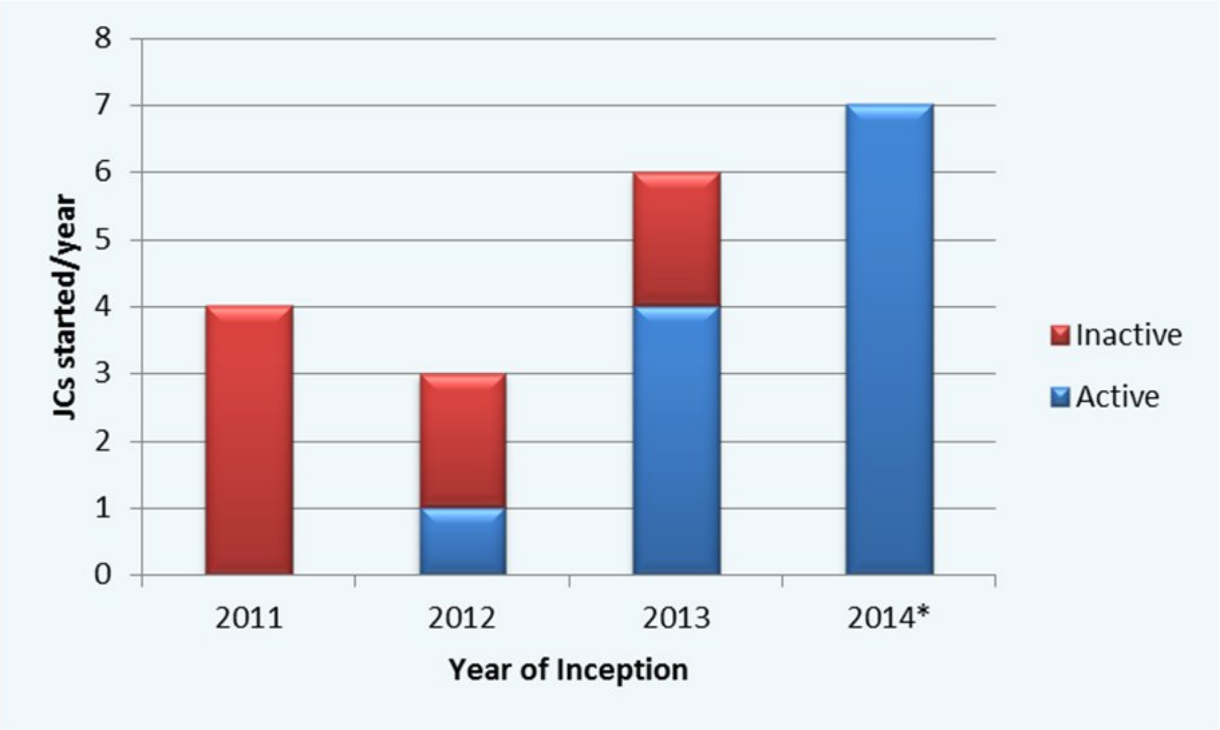
Hashtag	Associated Twitter user	Tweets	Followers	Total tweets	Mean tweet/mo[a]	Participants	Impressions/mo[a]
#1carejc[b]	@amcunningham	NA	NA	564	41	171	191,161
#ACCJournalclub[b]	@ACCinTouch	NA	NA	NA	NA	NA	NA
#ADC_JC	@ADC_JC	1452	459	2785	546	142	378,763
#ALiEMJC	@AnnalsofEM	NA	NA	924	80	234	136,145
#ambjc[b]	@ambjournalclub	57	87	93	3	42	1380
#bluejc	@bluejchost	516	202	3705	194	290	126,455
#cpjc[b]	@clinpsy*JC*	67	201	216	5.4	61	18,039
#ebnjc	@EBNursingBMJ	2117	1399	3901	395	456	719,241
#GeriMedJC	@gerimedJC	130	158	318	80	52	36,044
#HEJC	twubs/HEJC	NR	NR	986	38	103	18,176
#hpmjc	@hpmjc	588	129	1694	212	176	223,686
#IGSJC	@igsjc	624	750	5199	586	430	843,358
#jamapedsjc	@jamapeds	NR	NR	387	140	76	535,852
#JC_StE[b]	@JC_StE	615	374	1008	84	73	58,346
#MedEd[c]	@JournalGIM	NA	NA	NA	NA	NA	NA
#microtwjc	@microtwjc	525	155	NA	NA	NA	NA
#NephJC	@Nephjc	1436	584	5295	832	478	1,184,105
#PGHANJC	@BSPGHAN	NA	NA	NA	NA	NA	NA
#PHTwit*JC*	@PHTwit*JC*	1817	1057	4245	170	320	139,916
#psychjc	@psychiatryjc	92	72	240	218	44	288,109
#rsjc	@respandsleepjc	929	176	1669	401	115	86,730
#swjcchat[b]	@swjcchat	758	481	1199	138	138	6144
#twitjc[b]	@twitjournalclub	1498	3446	12,628	407	1,954	883,543
#urojc	@iurojc	1832	2401	9040	362	1,567	622,139

[a] During active period only.

[b] Inactive journal clubs.

[c] #MedEd is not a unique hashtag for this journal club—it is also used for discussion among other medical educators.

Figure 2. Establishment of journal clubs per year, comparing all journal clubs (blue) with currently active journal clubs (red). 2014 included journal clubs started prior to October 2014.



Determinants of Journal Club Performance

Tweets and impressions for each of the top five journal clubs as determined by tweets/impressions per month (ie, @ADC_JC, @EBNursingBMJ, @igsjc, @iurojc and @NephJC) were analyzed. Subgroups were created based on the "Top 10" participants for each month (as determined by Symplur) versus the remaining participants. An average of two tweets per month, was estimated for participants outside of the "Top 10" tweet contributors for each month (Figure 3). The contributions of the "Top 10" to overall tweets were significantly lower for #urojc (*P*<.005). Of those appearing in the "Top 10" for each month during journal club discussions, the majority were classified in the "Top 10" for the first or second time.

In an attempt to measure "reach" for each journal club, while considering large tweet traffic from these accounts in moderating and advertising, an impression:tweet ratio was calculated for each journal club user account. This impression:tweet ratio was shown to be linearly increasing for all journal clubs (R^2>.87 for all; Figure 4).

Figure 3. Average tweets per participant for the top 5 journal clubs (#ebnjc, #iurojc, #Nephjc, #ADC_JC, #igsjc). The overall calculated average for each journal club is represented by solid lines, while the average for participants outside of the Top 10 is represented by dashed lines.

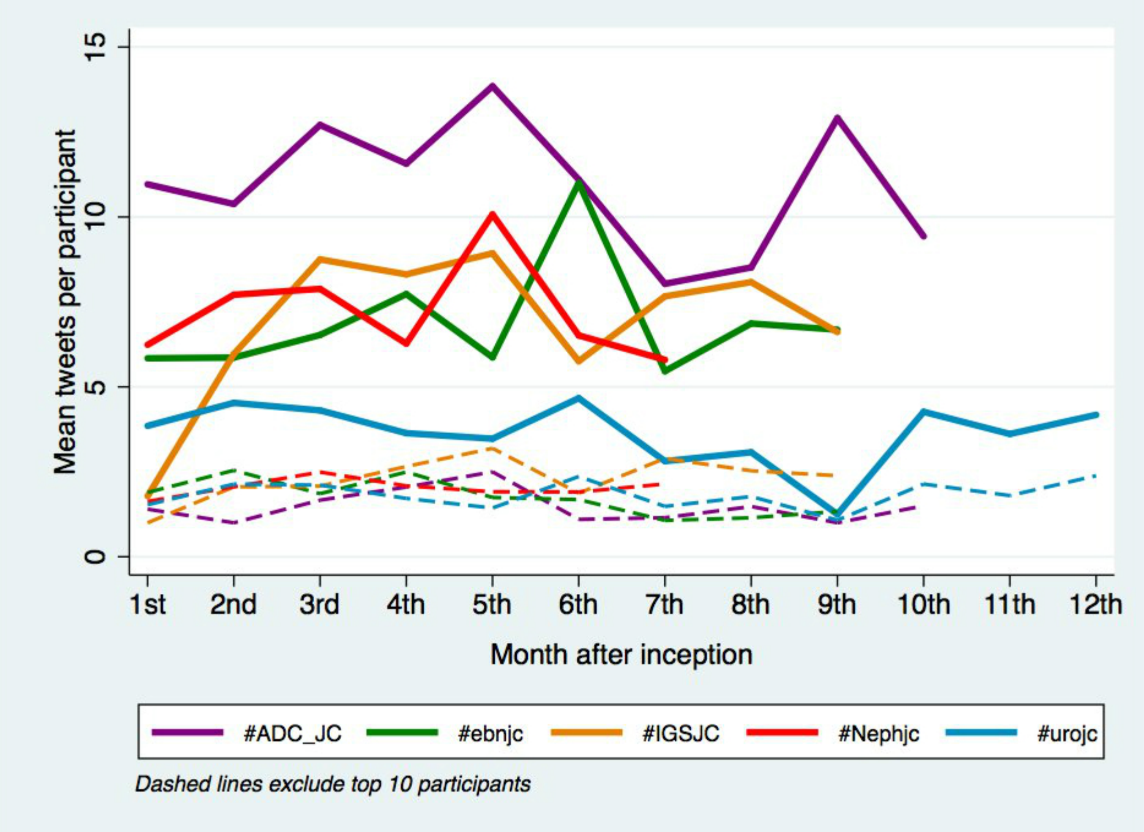
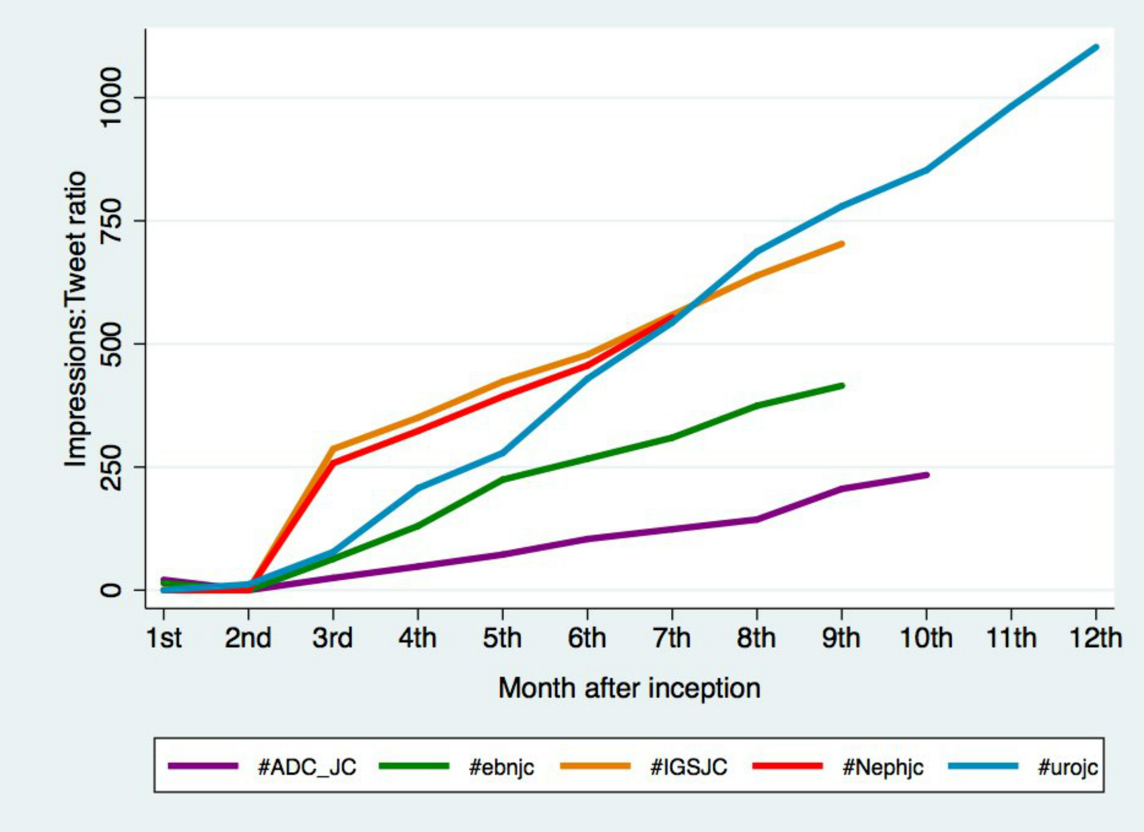


Figure 4. Journal club user account impressions:tweets ratio as a measure of reach, represented as absolute ratio change per month.



Discussion

Principal Findings

There were an estimated 200 million users per month actively tweeting an average of 500 million times per day in 2013, depicting Twitter as a contemporary, ever-changing social media environment. While recent reports in media outlets suggest that the rate of Twitter use overall may be declining, the findings of this systematic review of published literature and Twitter suggest that Twitter-based journal clubs are an expanding method of continuing professional development and a platform for global interaction. Published reports describe positive initial uptake and support from their respective medical communities [11,16-18], while Twitter analytics confirm an increasing trend in journal club establishment, cumulative tweets, and tweets per month.

Overall, we have observed an increasing participation in Twitter by the medical and allied health community, evidenced by an increasing year-to-year establishment of new journal clubs (Figure 2) and exponential increase in cumulative tweets and impressions. While the increase in tweets and impressions may be a function of journal clubs being established, it is more likely reflective of an ever-increasing global audience, reflected by the increasing impressions:tweets ratio for all journal clubs analyzed (Figure 4), and involvement of a wider, global audience viewing and participating in these discussions. Although the benefits of international involvement cannot be overemphasized, the benefits to local and regional formats may also be beneficial in communicating best practice or regional guidelines for countries and health jurisdictions where geographic separation is a significant obstacle. Given the current information explosion in medical research, Twitter also represents a potentially credible alternative to traditional “Commentary” pieces in peer-reviewed journals, allowing input from multiple key opinion leaders not previously available.

Similar increases in Twitter participation have been observed for medical conferences [7] and established peer-reviewed journals [25]. This increasing involvement by the global health audience appears to benefit traditional peer-review establishments. For journals, this increase in audience exposure appears to be beneficial for performance indices, with Twitter profiles associated with a higher mean impact factor for journals [25]. Furthermore, tweets have been reported to be predictive of future citations, with highly tweeted articles 11 times more likely to be highly cited than less tweeted articles in one journal [26].

While this relationship between social media and traditional academic media continues to grow, the use of Twitter for continuing professional development is an attractive venture. However, the freedom of voluntary participation complicates the establishment of an accurate and efficient record of participation for appropriate ethical acknowledgement for continuing professional development requirements by credentialing authorities. To date, no objective evaluation assessing the knowledge uptake and retention resulting from microblogging journal club is available [11]. Some strategies could include a posttest, similar to that provided by peer-reviewed journals, or a survey distributed by Twitter and completed online. These strategies may also be valuable in assessing participation in Free Open Access Medical Education (FOAM), which has also been facilitated by social media (eg, @UrologyQuiz found on Twitter).

FOAM encapsulates a collection of resources and tools for learning in medicine that are transforming medical education [9]. FOAM has continued to expand using social media as different collaborative teaching resources in accordance with popularity of online digital media, including blogs, podcasts, tweets, Google hangouts, online videos, and Facebook groups [27]. Twitter-based journal clubs are another addition to the FOAM sphere, available on Twitter (using hashtag #FOAMed). These resources allow individuals to upload medical content with discussion, collaboration, and dissemination of knowledge among users occurring on the individual site. The FOAM movement is a contemporary approach to improving and adding new collaborative resources to Web-based medical/health education, continuing professional development and research services [28]. Critics of FOAM have suggested that because such resources can be easily published online without quality control mechanisms, unreviewed FOAM resources may be erroneous or biased [29]. Peer-review processes for FOAM publications have been recently implemented in the form of high subject expertise from clinicians as either a pre-publication review and linked blog post for further comments, or post-publication review and facility for the author to amend or respond to the expert comments [29]. The Social Media Index (SM-i) is an emerging comparative index tool for FOAM resources, which combines quantitative online data to provide an overall rank to be calculated for any FOAM site [30], similar to other established peer-reviewed instruments comparing scholars (h-index), journals (Impact Factor), and websites (HONcode, DISCERN).

When individual journal club performance was considered, we observed a clear increase in overall tweets and impressions for established journal clubs, with some increasing at a rate as high as 722 tweets/month. However, month-to-month tweeting was observed to be highly variable. This variability could relate to diversity in interest among followers regarding the topics discussed or reduced participation of influential or high-volume tweeters for various reasons. There may also be limitations in access to articles discussed, particularly for individuals without institutional or individual journal access for those articles that are not open access. Where possible, some journal clubs, such as #urojc, provide open access to the discussed articles in order to remove this limitation [11].

In considering the determinants of journal club performance, subgroup analysis suggested that a large proportion of tweets each month were from those in the “Top 10”, with the remaining participants expected to contribute two tweets per journal club discussion (Figure 3). However, the majority of users appearing in the “Top 10” for each month during journal club discussions were doing so for the first or second time. This observation suggests that the broader audience participation and occasional generation of high traffic by eager participants outweighs the influence of regularly active tweeters or key Twitter-opinion leaders, who promoted the journal clubs to gain an initial following. However, in this open environment, there is also potential to overload the discussion or self-promote, creating unnecessary "noise". This focus on gaining followers and impressions as well as "mentions", which are similar to citations of a peer-reviewed manuscript may be strategic in improving a user's Klout score, an overall social media popularity rating [25]. Furthermore, the significantly lower contribution from the "Top 10" for #urojc compared to other journal clubs may be a consequence of longer establishment of #urojc, thus gaining a broader contribution to the discussion. The observed increasing impressions:tweets ratio suggested that the reach of the journal clubs analyzed continues to grow with time, resulting an in ever-expanding audience for these high-level academic discussions.

Limitations

This study is an analysis of social media at a point in time, when social media is known to be ever evolving. Much of the Twitter-based analytics relied on third-party services, such as Symplur, which was intended to focus on social media related to health care. Further, there are inherent limitations with the use of Twitter-based outcome measures, such as impressions, as a surrogate for reach. We were unable to measure the pattern of growth in followers or participants through data acquisition restrictions. There is also no current way to measure the passive value of journal clubs, specifically relating to users who are following the journal club discussion, and thus acquiring educational value, but not actively participating. We acknowledge that the number of included manuscripts is small, reflecting the current state of published literature relating to Twitter-based journal clubs. This review will serve as a check point and reference for the development of enhanced Twitter-based journal clubs by other medical craft groups internationally or loco-regionally.

Conclusions

This systematic review provides an illustration of early trends in the development of journal clubs using Twitter as a communication medium. Twitter-based journal clubs provide access to free, time-efficient, and high-level discussions on clinically important issues and equal participation opportunity for users. Twitter provides an unprecedented method of networking and formation of friendships with colleagues, which can be harnessed to educate, initiate research collaborations, and even canvass opinions with difficult cases in the time between conferences. Hence, the role of social media in continuing professional development will continue to evolve with increasing engagement by journals, conferences, and FOAM sources. Furthermore, in the midst of busy clinical duties, microblogging using Twitter provides a unique pathway to access and engage in discussions with peers and professional leaders.

Acknowledgments

MJR is supported by a Doctor in Training Research Scholarship from Avant Mutual Group Ltd., Cancer Council Queensland PhD Scholarship, and Professor William Burnett Research Fellowship from the Discipline of Surgery, School of Medicine, University of Queensland.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Terms used for search of literature databases.

[PDF File (Adobe PDF File), 79KB - jmir_v17i4e103_app1.pdf]

References

1. Leung EYL, Tirlapur SA, Siassakos D, Khan KS. #BlueJC: BJOG and Katherine Twining Network collaborate to facilitate post-publication peer review and enhance research literacy via a Twitter journal club. BJOG 2013 May;120(6):657-660. [doi: 10.1111/1471-0528.12197] [Medline: 23565946]
2. Linzer M. The journal club and medical education: over one hundred years of unrecorded history. Postgrad Med J 1987 Jun;63(740):475-478 [FREE Full text] [Medline: 3324090]
3. Sidorov J. How are internal medicine residency journal clubs organized, and what makes them successful? Arch Intern Med 1995 Jun 12;155(11):1193-1197. [Medline: 7763125]
4. Alguire PC. A review of journal clubs in postgraduate medical education. J Gen Intern Med 1998 May;13(5):347-353 [FREE Full text] [Medline: 9613892]
5. Deenadayalan Y, Grimmer-Somers K, Prior M, Kumar S. How to run an effective journal club: a systematic review. J Eval Clin Pract 2008 Oct;14(5):898-911. [doi: 10.1111/j.1365-2753.2008.01050.x] [Medline: 19018924]
6. Millichap JJ, Goldstein JL. Neurology Journal Club: a new subsection. Neurology 2011 Aug 30;77(9):915-917. [doi: 10.1212/WNL.0b013e31822b00f0] [Medline: 21876198]

7. Wilkinson SE, Basto MY, Perovic G, Lawrentschuk N, Murphy DG. The social media revolution is changing the conference experience: analytics and trends from eight international meetings. BJU Int 2015 May;115(5):839-846. [doi: 10.1111/bju.12910] [Medline: 25130687]
8. Grajales 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 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
9. Nickson CP, Cadogan MD. Free Open Access Medical education (FOAM) for the emergency physician. Emerg Med Australas 2014 Feb;26(1):76-83. [doi: 10.1111/1742-6723.12191] [Medline: 24495067]
10. Macrae HM, Regehr G, McKenzie M, Henteleff H, Taylor M, Barkun J, et al. Teaching practicing surgeons critical appraisal skills with an Internet-based journal club: A randomized, controlled trial. Surgery 2004 Sep;136(3):641-646. [doi: 10.1016/j.surg.2004.02.003] [Medline: 15349113]
11. Thangasamy IA, Leveridge M, Davies BJ, Finelli A, Stork B, Woo HH. International Urology Journal Club via Twitter: 12-month experience. Eur Urol 2014 Jul;66(1):112-117. [doi: 10.1016/j.eururo.2014.01.034] [Medline: 24548686]
12. Bowles PFD, Marenah K, Ricketts DM, Rogers BA. How to prepare for and present at a journal club. Br J Hosp Med (Lond) 2013 Oct;74(10):C150-C152. [Medline: 24273765]
13. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
14. Roberts MJ, Williamson DA, Hadway P, Doi SAR, Gardiner RA, Paterson DL. Baseline prevalence of antimicrobial resistance and subsequent infection following prostate biopsy using empirical or altered prophylaxis: A bias-adjusted meta-analysis. Int J Antimicrob Agents 2014 Apr;43(4):301-309. [doi: 10.1016/j.ijantimicag.2014.01.008] [Medline: 24630305]
15. Roberts M, Brodribb W, Mitchell G. Reducing the pain: a systematic review of postdischarge analgesia following elective orthopedic surgery. Pain Med 2012 May;13(5):711-727. [doi: 10.1111/j.1526-4637.2012.01359.x] [Medline: 22494470]
16. Chan TM, Rosenberg H, Lin M. Global emergency medicine journal club: social media responses to the January 2014 online emergency medicine journal club on subarachnoid hemorrhage. Ann Emerg Med 2014 Jul;64(1):88-94. [doi: 10.1016/j.annemergmed.2014.05.015] [Medline: 24951414]
17. Thoma B, Rolston D, Lin M. Global emergency medicine journal club: social media responses to the march 2014 annals of emergency medicine journal club on targeted temperature management. Ann Emerg Med 2014 Aug;64(2):207-212. [doi: 10.1016/j.annemergmed.2014.06.003] [Medline: 25059776]
18. Radecki RP, Rezaie SR, Lin M. Annals of Emergency Medicine Journal Club. Global Emergency Medicine Journal Club: Social media responses to the November 2013 Annals of Emergency Medicine Journal Club. Ann Emerg Med 2014 Apr;63(4):490-494. [doi: 10.1016/j.annemergmed.2014.01.033] [Medline: 24655449]
19. Leung EYL, Siassakos D, Waite L. Authors' reply: #BlueJC: BJOG and Katherine Twining Network collaborate to facilitate post-publication peer review and enhance research literacy via a Twitter journal club. BJOG 2013 Dec;120(13):1700. [doi: 10.1111/1471-0528.12452] [Medline: 24589003]
20. Lewis TL, Rallon P. #BlueJC: BJOG and Katherine Twining Network collaborate to facilitate post-publication peer review and enhance research literacy via a Twitter journal club. BJOG 2013 Dec;120(13):1699-1700. [doi: 10.1111/1471-0528.12451] [Medline: 24589002]
21. Leung EY, Malick SM, Khan KS, EBM-CONNECTCollaboration. On-the-Job Evidence-Based Medicine Training for Clinician-Scientists of the Next Generation. Clin Biochem Rev 2013 Aug;34(2):93-103 [FREE Full text] [Medline: 24151345]
22. Queensland Nurse. 2013. Medical Journal Adds Tweet Chat URL: http://www.readperiodicals.com/201304/2960920081.html [accessed 2015-04-21] [WebCite Cache ID 6XyADMoGw]
23. Linton KD, Woo HH, Twitter International Urology Journal Club. Complications of prostate cancer treatment. Lancet Oncol 2014 Apr;15(4):e150-e151. [doi: 10.1016/S1470-2045(14)70092-0] [Medline: 24694634]
24. Mehta N, Flickinger T. The times they are a-changin': academia, social media and the JGIM Twitter Journal Club. J Gen Intern Med 2014 Oct;29(10):1317-1318. [doi: 10.1007/s11606-014-2976-9] [Medline: 25092011]
25. Nason GJ, O'Kelly F, Kelly ME, Phelan N, Manecksha RP, Lawrentschuk N, et al. The emerging use of Twitter by urological journals. BJU Int 2015 Mar;115(3):486-490. [doi: 10.1111/bju.12840] [Medline: 24925047]
26. 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]
27. Cadogan M, Thoma B, Chan TM, Lin M. Free Open Access Meducation (FOAM): the rise of emergency medicine and critical care blogs and podcasts (2002-2013). Emerg Med J 2014 Oct;31(e1):e76-e77. [doi: 10.1136/emermed-2013-203502] [Medline: 24554447]
28. Boulos MNK, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. BMC Med Educ 2006;6:41 [FREE Full text] [doi: 10.1186/1472-6920-6-41] [Medline: 16911779]
29. Thoma B, Chan T, Desouza N, Lin M. Implementing peer review at an emergency medicine blog: bridging the gap between educators and clinical experts. CJEM 2014 Apr 1;16:1-4. [Medline: 24742824]

30. Thoma B. Social Media Index (SM-i). 2013. URL: http://brentthoma.mededlife.org/portfolio/social-media-index-sm/ [accessed 2015-01-01] [WebCite Cache ID 6VFKc4Pjt]

Abbreviations

FOAMed: Free Online Access Medical Education
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 01.01.15; peer-reviewed by T Chan, H Wu; comments to author 22.01.15; revised version received 21.02.15; accepted 23.02.15; published 23.04.15.

Please cite as:
Roberts MJ, Perera M, Lawrentschuk N, Romanic D, Papa N, Bolton D
Globalization of Continuing Professional Development by Journal Clubs via Microblogging: A Systematic Review
J Med Internet Res 2015;17(4):e103
URL: http://www.jmir.org/2015/4/e103/
doi:10.2196/jmir.4194
PMID:25908092

Original Paper

Eliciting and Receiving Online Support: Using Computer-Aided Content Analysis to Examine the Dynamics of Online Social Support

Yi-Chia Wang[1], MSc; Robert E Kraut[2], PhD; John M Levine[3], PhD

[1]Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, United States
[2]Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, United States
[3]Department of Psychology, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Yi-Chia Wang, MSc
Language Technologies Institute
School of Computer Science
Carnegie Mellon University
Language Technologies Institute
5000 Forbes Ave. Gates Hillman Complex 5404
Pittsburgh, PA, 15213
United States
Phone: 1 412 736 2538
Fax: 1 412 268 6298
Email: yichiaw@cs.cmu.edu

Abstract

Background: Although many people with serious diseases participate in online support communities, little research has investigated how participants elicit and provide social support on these sites.

Objective: The first goal was to propose and test a model of the dynamic process through which participants in online support communities elicit and provide emotional and informational support. The second was to demonstrate the value of computer coding of conversational data using machine learning techniques (1) by replicating results derived from human-coded data about how people elicit support and (2) by answering questions that are intractable with small samples of human-coded data, namely how exposure to different types of social support predicts continued participation in online support communities. The third was to provide a detailed description of these machine learning techniques to enable other researchers to perform large-scale data analysis in these communities.

Methods: Communication among approximately 90,000 registered users of an online cancer support community was analyzed. The corpus comprised 1,562,459 messages organized into 68,158 discussion threads. Amazon Mechanical Turk workers coded (1) 1000 thread-starting messages on 5 attributes (positive and negative emotional self-disclosure, positive and negative informational self-disclosure, questions) and (2) 1000 replies on emotional and informational support. Their judgments were used to train machine learning models that automatically estimated the amount of these 7 attributes in the messages. Across attributes, the average Pearson correlation between human-based judgments and computer-based judgments was .65.

Results: Part 1 used human-coded data to investigate relationships between (1) 4 kinds of self-disclosure and question asking in thread-starting posts and (2) the amount of emotional and informational support in the first reply. Self-disclosure about negative emotions (beta=.24, *P*<.001), negative events (beta=.25, *P*<.001), and positive events (beta=.10, *P*=.02) increased emotional support. However, asking questions depressed emotional support (beta=–.21, *P*<.001). In contrast, asking questions increased informational support (beta=.38, *P*<.001), whereas positive informational self-disclosure depressed it (beta=–.09, *P*=.003). Self-disclosure led to the perception of emotional needs, which elicited emotional support, whereas asking questions led to the perception of informational needs, which elicited informational support. Part 2 used machine-coded data to replicate these results. Part 3 analyzed the machine-coded data and showed that exposure to more emotional support predicted staying in the group longer 33% (hazard ratio=0.67, *P*<.001), whereas exposure to more informational support predicted leaving the group sooner (hazard ratio=1.05, *P*<.001).

Conclusions: Self-disclosure is effective in eliciting emotional support, whereas question asking is effective in eliciting informational support. Moreover, perceptions that people desire particular kinds of support influence the support they receive. Finally, the type of support people receive affects the likelihood of their staying in or leaving the group. These results demonstrate the utility of machine learning methods for investigating the dynamics of social support exchange in online support communities.

(J Med Internet Res 2015;17(4):e99) doi:10.2196/jmir.3558

KEYWORDS

social support; health communication; self-disclosure; social media; support groups; emotions; natural language processing

Introduction

Background

A high percentage of people with chronic or life-threatening diseases use online resources to obtain information about their condition and ways to cope with it. Although informational websites are the most popular, many people—especially cancer patients and survivors—participate in online health support communities [1,2]. A recent meta-analysis suggested that online support communities are effective in decreasing depression and increasing self-efficacy and quality of life [3]. Although several clinical trials suggest that participation in Internet-based support communities improves emotional well-being (eg, [4,5]), conclusions are ambiguous because most interventions have multiple components of which support group participation is only a part [6]. Moreover, research also shows that support interventions often do not provide the benefits they were designed to produce (eg, [7]). Thus, much remains to be learned about when and why support is effective in online communities.

Of the several categories of social support that have been identified in offline and online communities, emotional and informational support have received the most theoretical and empirical attention. Ridings and Gefen [8] reported that 76% of people who join online health communities do so to exchange emotional and informational support. Emotional support refers to the provision of caring, sympathy, or encouragement, whereas informational support refers to the provision of information or advice. Evidence in offline settings indicates that cancer patients seek out emotional support [9] and claim that it is especially helpful [10]. And research suggests that peer discussion focusing on emotional support enhances cancer patients' psychological adjustment [11,12]. Participants in cancer support communities also exchange information and advice about diagnoses, treatments, adverse effects, relations with physicians, financial problems, and so on. And research suggests that information exchanged in offline support groups is associated with improvement in psychological well-being [13].

Interestingly, most research on social support, both offline and online, has focused on its physical and psychological effects, rather than on how it is elicited. For this reason, very little is known about the strategies that people use to seek support from others. This is surprising because the exchange of support is a dynamic communication process involving actions by support seekers and support providers [14]. During support elicitation, support seekers use a variety of techniques to indicate their need for support. During support provision, support providers recognize seekers' behavior as requests for support and decide how to respond.

Strategies for Obtaining Social Support

The Social Support Activation Model [15] categorizes support-seeking behaviors along 2 dimensions: (1) direct versus indirect elicitation and (2) verbal versus nonverbal elicitation. In online environments, where communication is verbal, a direct elicitation strategy might involve asking a question about a medication, whereas an indirect strategy might involve expressing fear about a diagnosis. Moreover, there is suggestive evidence that these 2 kinds of strategies may be used to elicit different kinds of support. For example, prior research suggests that participants in online support communities ask questions to get factual information [16] and that focused questions are more effective in getting useful answers than are open-ended ones [17]. In contrast, to elicit emotional support, participants often tell narratives about their disease and their emotional reactions to it [16,18]. The goal of this research is to further clarify the process by which members of online cancer support communities obtain informational and emotional support from other members of their group.

An important feature of our research is its reliance on automated techniques for analyzing a large number of conversational exchanges between community members. Most studies on communication in support communities are based on hand coding relatively small samples of messages (eg, [16,18-22]). Even Meier and colleagues' [21] relatively ambitious effort only coded emotional and informational support in approximately 3000 messages. Because online support communities usually generate large numbers of posts, it is impractical to hand code all messages. Previous research has shown that it is feasible to partially automate some text analyses of conversations in online support communities (eg, [23,24]), but there has been little effort to fully automate the analysis of large numbers of messages in these communities.

Utilizing large-scale data and computer-aided content analysis is valuable for 3 reasons. First, because machine learning models can usually be easily replicated, they enable researchers interested in similar research questions to deploy comparable methods to challenge, verify, or extend others' results. Second, the large-scale data that can be analyzed by machine learning methods allow researchers to answer more subtle research questions, conduct finer-grained analyses, and examine longer-term interaction patterns among participants. For example, using data from almost 60,000 exchanges from more than 30,000 participants in a breast cancer support community over a 9-year period, we have shown that being exposed to emotional and informational support has different effects on members' subsequent participation in the group [25]. Third, computer-aided content analysis opens up opportunities for real-time interventions. For example, an online support community can use an automated model to detect the type of support an author of a particular post is seeking and then direct that person to discussions or other participants most relevant to his or her need.

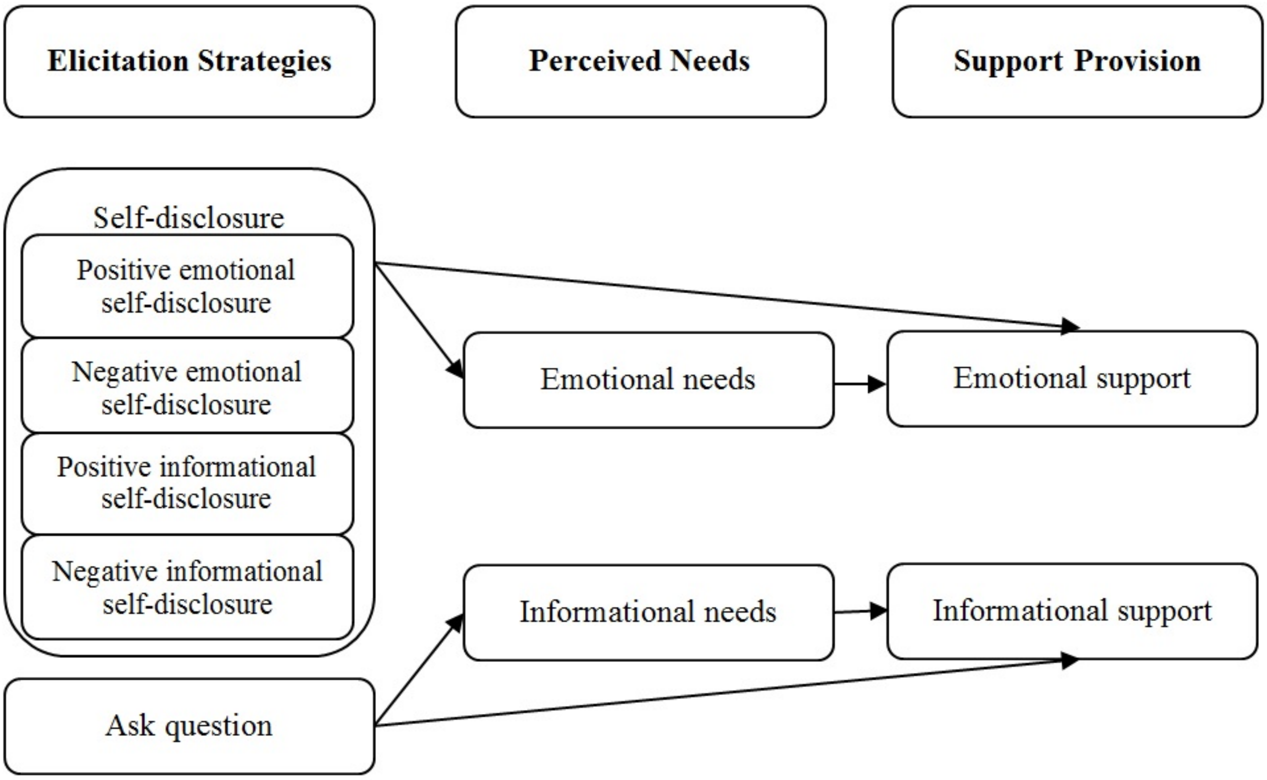
In this paper, we use archival data to examine the relationships between the conversational moves people make in thread-starting posts and the amount and type of support they receive in response to them. In part 1, we conducted the analysis using a relatively small sample of approximately 1000 hand-coded conversational threads. In part 2, we replicated these results using machine learning algorithms to analyze more than 67,000 conversational threads. These algorithms use statistical procedures analogous to multiple regression to correlate language characteristics of messages with human judgments of them [26]. Specifically, part 2 describes techniques to automatically identify the extent to which messages exchanged in breast cancer discussion forums sought and provided emotional and informational support. Parts 1 and 2 test hypotheses suggested by, but not rigorously tested in, prior studies: (1) self-disclosure of support seekers elicits emotional support, whereas (2) asking questions elicits informational support. Finally, part 3 demonstrates additional benefits of automated coding by briefly reviewing previously published analyses showing that the type of support participants are exposed to predicts their continued participation in the group.

Conceptual Model

Figure 1 shows the model underlying our research. The model posits that different elicitation strategies on the part of support seekers (personal self-disclosure vs asking questions) produce different kinds of support from providers (emotional support vs informational support, respectively). Moreover, the model posits that this effect is mediated by the perceived needs of the support seeker. More specifically, the model posits that self-disclosure, irrespective of its content (emotional, informational) or valence (positive, negative), leads to perceived emotional needs, which in turn elicit emotional support. In contrast, asking a question leads to perceived informational needs, which in turn elicit informational support.

In this model, we hypothesize that the effects of elicitation strategies on support are mediated by the perceived needs of the support seeker. This hypothesis is important because, without such a process, a relationship between a seeker's behavior and others' responses could be interpreted as behavioral mimicry. Behavioral mimicry is a well-established phenomenon in human communication [27]; it occurs for both language (eg, [28]) and nonverbal behavior (eg, [29]), and it leads to better language understanding and interpersonal rapport. In conversations in support groups, behavioral mimicry could be confused with attempts to meet perceived needs. For example, most types of self-disclosure use first-person singular pronouns in constructions such as, "Well it's been 2 weeks since my doctor called and gave me the results and I have been up and down since" or "I'm a 35 y/o happily married mother of 2 and completely and utterly shaken to my core." People who give emotional support by replying with their own cancer experiences (eg, "I remember being in your place. I am only 4 months out from bilateral mastectomy and things have improved tremendously. Life will get better.") are also likely to use first-person pronouns.

Figure 1. Conceptual model of social support elicitation and provision.



Part 1: Using Human-Coded Data to Model How People Elicit Emotional and Informational Support

Overview

Part 1 investigated the language strategies that support group members use to elicit emotional and informational support and the meditational role of their perceived needs. Because each of the constructs in Figure 1 is based on human judgments, this part of the study was limited to an analysis of approximately 1000 exchanges.

Methods

Research Site

We studied support conversations in a large online breast cancer support community with a variety of communication platforms, including discussion boards and chat rooms. The discussion board platform was one of the most popular and active online breast cancer support communities on the Internet. At the time of our research, it had more than 90,000 registered members and 66 forums organized by criteria such as disease stage (eg, metastatic breast cancer), treatment (eg, hormonal therapy), demographic characteristics (eg, women aged 40-60 years), and treatment options (eg, breast reconstruction). In the forums, members ask questions, share stories, and read posts about how to deal with their disease. This discussion board platform is a rich environment for studying the dynamics of online support communities.

We collected all public posts on the discussion boards of the online support community from October 2001 to January 2011. During this period, the boards contained a total of 1,562,459 messages in 68,158 discussion threads. The median length of a discussion thread was 6 messages (mean 22.9, SD 280.5). The median life span of a thread, from the first thread-starting message to the last, was 2 days (mean 29.3, SD 115.6). A total of 81.12% (55,291/68,158) of thread-starting messages received a response within 24 hours, whereas 11.42% (7785/68,158) never received a response.

Measuring Elicitation Strategies, Perceived Support Needs, and Support Provision

We employed Amazon Mechanical Turk (MTurk) workers to judge how many messages from a random sample of 1000 thread starters and their first replies contained each of the constructs in Figure 1, such as negative informational self-disclosure or emotional support. MTurk is an online marketplace for crowdsourcing. It allows requesters to post jobs, called Human Intelligence Tasks (HITs), which workers, known as Turkers, can choose to perform. Snow et al [30] have shown that the combined judgments of a small number (between 5 to 7) of naïve judges on MTurk lead to ratings of texts that are very similar to those that experts make on such dimensions as emotions expressed, relative timing of events referred to, word similarity, word sense disambiguation, and linguistic entailment or implication.

We created independent tasks for judging each concept in Figure 1. One group of Turkers rated how many thread starter messages contained self-disclosure about events in the poster’s life or her thoughts and feelings. A second group rated how many thread starters asked questions. A third group rated how many thread starters sought emotional or informational support. Finally, a fourth group was shown the first replies to the thread starters (with the corresponding thread starters for reference) and judged how many provided informational and emotional support. The wording for the 9 judgments Turkers made are listed in Table 1. The Turkers made their judgments using a 7-point Likert scale (1=none; 7=a great deal).

Table 1. The tasks Turkers performed and the resulting intraclass correlation (ICC) for each construct.

Construct	ICC	Definition for Turkers	Prompt for Turkers
Emotional self-disclosure		Emotional self-disclosure is concerned with the extent to which the writer has discussed her feelings and emotions with others, such as happiness, fears, sadness, and anger.	
Positive emotional self-disclosure	0.90	Example of positive emotional self-disclosure: “Now that chemo is done, I find myself waking up in the morning feeling a huge burden has been lifted from my shoulders.”	To what extent does this message contain positive emotional self-disclosure?
Negative emotional self-disclosure	0.94	Example of negative emotional self-disclosure: “I am freaked out after reading my mammogram report.”	To what extent does this message contain negative emotional self-disclosure?
Informational self-disclosure		Informational self-disclosure is concerned with the extent to which the writer has discussed her personal information with others, such as health conditions, diagnosis results, and family status. Informational self-disclosure can be related to the positive, negative, or neutral life events of the writer.	
Positive informational self-disclosure	0.85	Example of positive informational self-disclosure: “Took family to Cleveland Zoo for the first time in years and years.”	To what extent does this message contain positive informational self-disclosure?
Negative informational self-disclosure	0.91	Example of negative informational self-disclosure: “I found a lump in my armpit about 5 weeks ago. It’s not fixed, but moveable. I have periodic tingling or single sharp pains in my left breast every once in a while.”	To what extent does this message contain negative informational self-disclosure?
Asking a question	0.91	When asking a question, the writer is requesting a response from the group. Questions can be asked directly and indirectly. Examples of questions: “What will you go through for a small chance to live longer?” “So I guess my question is, can a NOT dense breast, just average, have missed tumors on imaging?” “I am wondering if anyone has any advice on what I should do?” “Looking for any insight others may have on this.”	To what extent is this message asking a question?
Eliciting support			
Emotional support elicitation	0.91	When seeking emotional support, the writer is trying to get understanding, encouragement, affirmation, sympathy, or caring.	To what extent is this message seeking emotional support?
Informational support elicitation	0.95	When seeking informational support, the writer is trying to get advice, referrals, or knowledge.	To what extent is this message seeking informational support?
Providing support		There are 2 kinds of social support: emotional support and informational support.	
Provide emotional support	0.92	Emotional support messages provide understanding, encouragement, affirmation, sympathy, or caring.	How much emotional support does this message provide?
Provide informational support	0.92	Informational support messages provide advice, referrals, or knowledge.	How much informational support does this message provide?

Ten Turkers made each judgment, with different subsets rating each message. We aggregated workers’ responses for each message by averaging their ratings. Turkers were paid US $0.03 per message for judging question asking and US $0.05 per message for judging emotional self-disclosure, informational self-disclosure, perceived support needs, and support provision. To encourage workers to take the numeric-rating task seriously, they also highlighted words and phrases in the message that provided evidence for their ratings. To further control the annotation quality, we restricted the worker pool to Turkers who indicated a United States location and had at least of 98% their work accepted by their previous Turk employers.

Table 2 provides descriptive statistics and correlations among the 7 constructs judged by the Turkers.

Table 2. Descriptive statistics and correlations among constructs coded by Turkers.

Variable	Mean (SD)	1	2	3	4	5	6	7	8
1. Positive emotional disclosure	1.55 (0.96)	1							
2. Negative emotional disclosure	2.39 (1.52)	–.06	1						
3. Positive informational disclosure	1.89 (1.09)	.76	–.09	1					
4. Negative informational disclosure	3.58 (1.72)	–.17	.68	–.24	1				
5. Question asking	4.94 (2.17)	–.35	.06	–.34	.31	1			
6. Emotional support elicitation	2.75 (1.66)	.13	.79	.09	.58	–.11	1		
7. Informational support elicitation	4.21 (2.01)	–.36	–.06	–.37	.26	.88	–.28	1	
8. Provide emotional support	2.68 (1.43)	.16	.39	.14	.32	–.17	.49	–.24	1
9. Provide informational support	2.93 (1.47)	–.23	–.01	–.23	.17	.42	–.13	.49	–.17

Following are 2 examples from our final hand-coded dataset. Example 1 shows a message with high emotional support and low informational support and example 2 shows a message with low emotional support and high informational support. The example messages are lightly disguised using the techniques suggested by Bruckman [31]. Although the examples illustrate cases in which one type of support is high and the other low, across the full hand-coded sample, the two types of support were only weakly negatively correlated (r_{972}=–.17, *P*<.001).

Example 1 (emotional support=5.7; informational support=1.0):

> *Julie-you have had such a difficult road, but yet you still manage to do well in school...I am truly inspired by you. Big cyber hugs and best wishes to you:>*

Example 2 (emotional support=1.2; informational support=4.5):

> *Extranodal extension occurs when the tumor extends through the wall of the lymph node. This is noted on pathology reports, but in the main it isn't very significant, and isn't used in assessing cancer stage.*

We assessed the reliability of raters' judgments using intraclass correlations (ICC), which indicate the proportion of the variance in judgments that can be attributed to the message being judged. Table 1 shows the ICCs for the social support-related constructs used in this research. We validated Turkers' judgments by comparing their judgments to expert judgments of the informational and emotional support contained in 50 messages selected from corpora studied by Meier et al [21] and Bambina [16]. The correlations between Turkers' and experts' average ratings for emotional support (*r*=.70) and informational support (*r*=.76) were both high.

Results

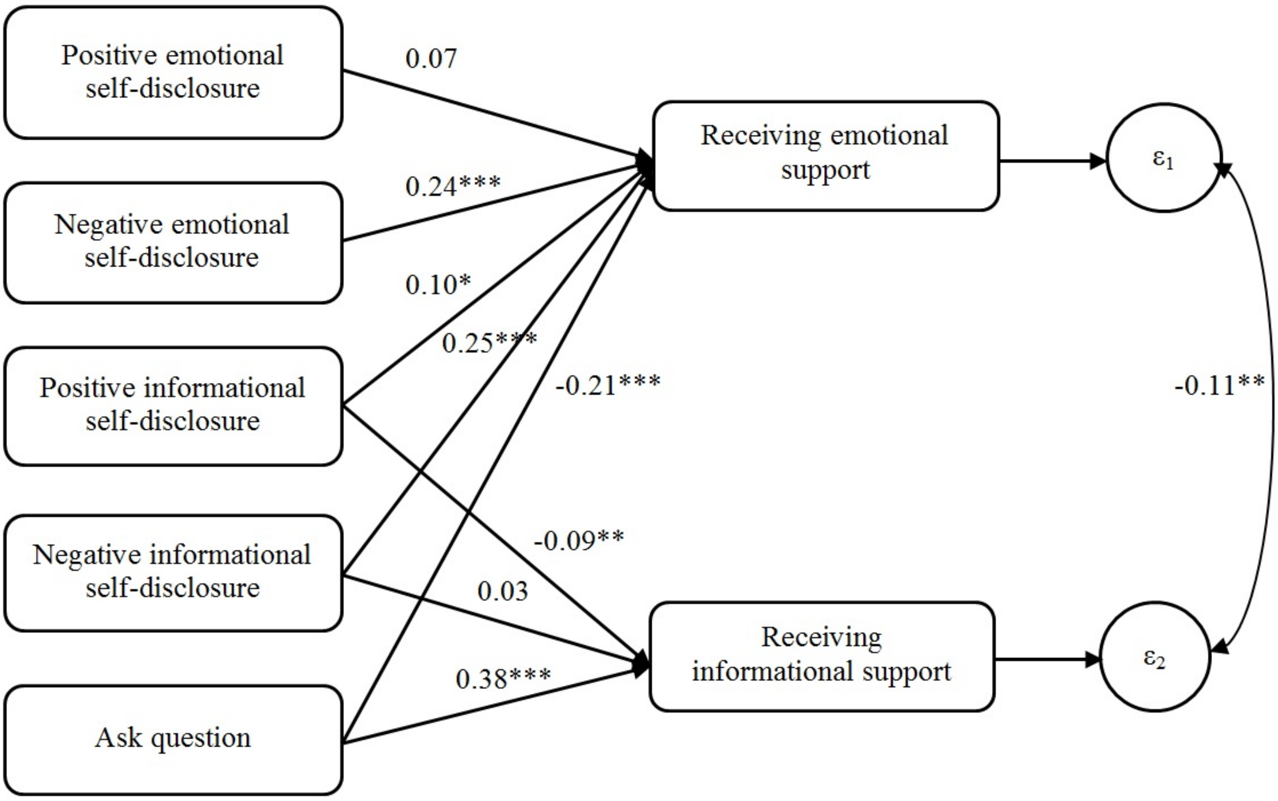
Relationship Between Elicitation Strategies and Support Provision

We first tested the hypotheses that self-disclosure elicits emotional support whereas questions elicit informational support using the dataset of 1000 Turker-coded messages. We used structural equation modeling to examine the relationships between the amount of self-disclosure and question asking in the thread starters and the amount of emotional and informational support provided in the first reply. Standardized regression coefficients (β) are reported. Figure 2 is the path model showing the results. This model fits the human-coded data well, with the comparative fit index (CFI)=0.98 and the root mean squared error of approximation (RMSEA)=0.062 (see [32] for rules-of-thumb for evaluating the goodness of fit for structural equation models; a CFI>0.97 is generally considered a good fit, whereas a RMSEA between 0.05 and 0.08 is generally considered an acceptable fit). Note that we did not include the relationships between positive/negative emotional self-disclosure and informational support provision in the model because these relationships were not significant and excluding them improved the model fit. This applied to all the models reported in this paper.

The more thread-starting messages contained negative emotional self-disclosure (beta=.24, SE 0.04, *P*<.001) and negative informational self-disclosure (beta=.25, SE 0.04, *P*<.001), the more responses to them provided emotional support. Positive informational self-disclosure was also associated with more emotional support in replies, although not as strongly (beta=.10, SE 0.04, *P*=.02). However, positive emotional self-disclosure was not significantly associated with more emotional support (beta=.07, SE 0.04, *P*=.10). In contrast to these effects of self-disclosure which generally increased emotional support, asking questions was associated with receiving less emotional support (beta=–.21, SE 0.03, *P*<.001).

Participants used different strategies to elicit informational support. Thread starters were more likely to get information and advice when they explicitly asked for it, but less likely to receive it when they described positive events in their lives. The more the thread starters asked questions, the more the first reply provided informational support (beta=.38, SE 0.03, *P*<.001). In contrast, when the thread starter revealed more positive informational self-disclosure, the reply contained less informational support (beta=–.09, SE 0.03, *P*=.003). Negative informational self-disclosure was not associated with receiving informational support (beta=.03, SE .03, *P*=.306). These results are consistent with prior research indicating that asking explicit questions and providing a rationale elicited information and advice [33,34]. In contrast, when thread starters described positive events in their lives, they were less likely to receive information and advice, even if they asked questions. Perhaps in these cases, recipients believed that the thread starter did not really need their help, although asking for it.

Figure 2. Path model showing the analysis of the social support communication process based on Turker-coded data. Values represent standardized regression coefficients. ε1 and ε2 represent error terms. * *P*<05; ** *P*<01; *** *P*<001.



Mediation Effect of Perceived Support Needs

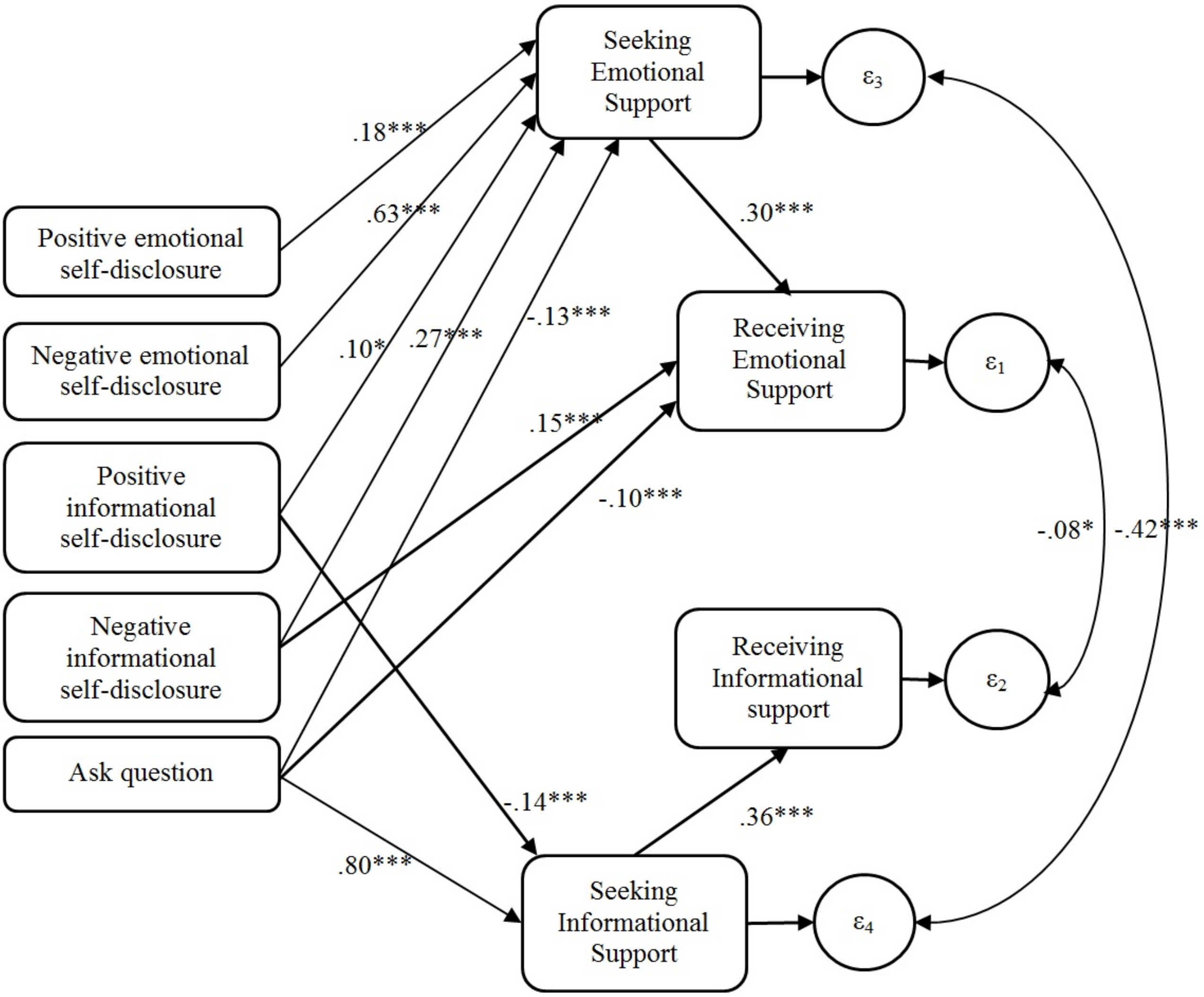
When people started a thread with self-disclosure and especially when they revealed negative thoughts and feelings and negative events in their lives, and when they refrained from asking questions, others were likely to offer them emotional support. However, when they asked questions and refrained from describing positive events in their lives, others offered them information and advice instead. We assumed these language features lead to social support because other participants in the community perceived them as indicators of a need for emotional support and informational support, respectively. That is, other members of the community treated these language features as requests for particular kinds of support. To test this assumption, we conducted mediation analysis to assess whether the language features in thread-starting messages have their effects on the emotional and information support in replies because they signaled a desire for a particular kind of support.

Results are shown in Figure 3, which includes only the direct paths between constructs. Although the CFI index (0.98) indicated that the mediation model was a good fit to the data, the RMSEA criterion (0.119) was problematic. As Schermelleh et al [32] noted, “it is quite difficult to decide on data-model fit or misfit, especially if various measures of model fit point to conflicting conclusions about the extent to which the model actually matches the observed data...[Although rules-of-thumb exist about what constitutes a good fitting SEM model], these rule-of-thumb cutoff criteria are quite arbitrary and should not be taken too seriously.” We interpret our findings as indicating mediation. As Figure 3 illustrates, people received emotional support when they were perceived as seeking it (beta=.30, SE 0.04, *P*<.001) and informational support when they were perceived as seeking that (beta=.36, SE 0.04, *P*<.001). The effects of the language features that predicted receiving either informational or emotional support in Figure 2 were all at least partially mediated by the perception that the author was seeking this particular kind of support, as shown in Figure 3. The data came from human judgments of each construct. Numbers represent standardized regression coefficients from a structural equation model. Only direct links between constructs significant at *P*<.05 are shown. Indirect effects can be calculated by multiplying the direct effects (eg, the indirect effects of asking questions on receiving informational support mediated by seeking informational support is .80*.36=.29).

Focusing on the receipt of emotional support, the effects of all the speech acts were either partially or completely mediated by their influence on judgments that the writer was seeking emotional support. Negative emotional self-disclosure (writing about sadness, fear, anger, and other negative thoughts and feelings) was associated with receiving more emotional support, and this effect was completely mediated by the perception that the writer was seeking emotional support (indirect effect/total effect=.19/.21=90%); the direct effect of negative emotional self-disclosure on getting emotional support became insignificant after controlling for the perception that the writer was seeking emotional support (beta=.02, *P*=.74). Similarly, the effect of positive informational self-disclosure on receiving emotional support was partially mediated by judgments that the writer was seeking emotional support (.03/.13=23%), leaving a nonsignificant direct path (beta=.10, *P*=.07). The effect of negative informational self-disclosure on receiving emotional support was also partially mediated by judgments that the writer was seeking emotional support (.08/.23=35%). The direct effect of negative informational self-disclosure on receiving emotional support decreased from .25 (in Figure 2) to .15 (*P*<.001) when controlling for the perception that the writer was seeking emotional support. The negative effect of asking questions on receiving emotional support was also partially mediated by the judgment that the writer was not seeking emotional support (–.04/–.14=29%). However, asking questions also had a direct negative effect on receiving emotional support (beta=–.10, *P*<.001).

Focusing on the receipt of informational support, the effects of positive and negative informational self-disclosure and asking questions were completely mediated by their direct influence on judgments that the writer was seeking informational support. Although people were likely to receive informational support after asking questions (total effect=.26), the total effect was mediated by the link between asking questions and seeking information and between seeking information and receiving it (.29/.26=112%). Similarly, 42% of the total negative association of positive information disclosure on receiving informational support was mediated by its association with the perception the writer was seeking informational support (–.05/–.12=42%).

Figure 3. Direct effects of language features in a thread-starting post on the perception that the poster was seeking emotional and informational support and on the receipt of emotional and informational support. Values represent standardized regression coefficients. ε1, ε2, ε3, and ε4 indicate error terms. * *P*<.05; ** *P*<.01; *** *P*<.001.



Part 2: Using Machine-Coded Data to Model How People Elicit Emotional and Informational Support

Methods

Overview

Part 2 sought to partially replicate the analyses previously described, which were based on human coding of 1000 interactions, by using machine learning techniques to automate the coding of 58,357 interactions. We constructed machine learning models to automatically measure both the language elicitation strategies posters used in their thread-starting posts and the emotional and informational support they received in the first reply to them. Our goal was to use machine learning approaches to correlate characteristics of messages with human judgments made about various elicitation strategies and the presence of emotional and informational support. As described subsequently, the input features included language characteristics of the messages, such as their length, presence of words from general and domain-specific dictionaries, syntax features (eg, the ordering of noun and verb phrases), higher-level semantic features (eg, the presence of advice and questions), and questions. The outputs were numerical values representing the extent to which a message asked questions or provided various types of self-disclosure or support. A model was successful if (1) its assessments of the outcomes of interest matched those produced by human judges and (2) it was parsimonious, using a small number of input features while producing an accurate assessment.

Building, validating, and applying the machine learning models involved 5 steps. First, human judges hand-coded language strategies and support provisions in a sample of messages. Their judgments represented the “ground truth” or “gold standard” to which we compared the machine learning estimates. Second, we represented the messages as a set of language features that could be interpreted by computers and would be the input to the machine learning algorithms. Third, we identified the best machine learning models from a part of the hand-coded data by iteratively testing different configurations of machine learning algorithms and combinations of input features. Fourth, to evaluate the performance of the final models, we applied them to a holdout sample of data. Finally, we tested hypotheses by applying the machine learning models to automatically code the 58,357 thread starter–first reply pairs in the entire dataset.

Building Machine Learning Models for Social Support Concepts

Because most health researchers will not be familiar with the methods used to construct the machine learning models to measure self-disclosure, question asking, and emotional and informational support, we will describe them and their rationale in some detail. For a general treatment of machine learning, we refer the reader to Witten et al [35]. The processes for building models for each construct in our model were similar, so we illustrate the general approach by describing how to measure the presence of emotional and informational support.

Step 1: Creating the Human-Coded Dataset

We used the annotations collected in part 1 as the human-coded dataset for machine learning model training. The details regarding how we obtained the human annotations were described in part 1.

Step 2: Identifying Language Features

In the second step, messages were represented as lower-level, computer-readable language features, which were likely to provide cues that a text was self-disclosing, asking a question, or offering emotional or informational support. We focused our search for language features in domains that we believed would reflect the topics and writing styles people use when they talk about their disease and their relationship to it. We focused on the 6 language domains:

1. Body: one of the most common conversational themes in online health support communities involves patients’ bodies and how disease changes their bodies. Participants might describe their bodily conditions to look for advice or to share their experiences with others. Sometimes they tell others the results of their treatment by mentioning changes in their bodies, such as weight loss or the size of a tumor. The discussion of the body can signal the offering of support and not just self-disclosure.
2. Positive/negative emotions and events: revealing underlying emotions or sharing personal life events and stories is a way to self-disclose and provide support. Online support participants often share personal feelings about themselves and use language that indicates happiness, fear, anger, and other emotional states. People also often describe positive things that happened in their lives, such as successful treatment, to encourage others and provide support to them.
3. Social: social relations are crucial for cancer survivors. Their disease may change how they interact with others. Through interaction with others, they can feel supported and gain the strength to face their disease. In the absence of support, they may feel abandoned.
4. Spiritual: spiritual or religious comments can provide social support, especially for those with strong religious beliefs. Examples of spiritual comments in this online breast cancer community are “God bless” and “I’ll add you to my prayer list.”
5. Time: time-related information is often mentioned in online support discussions. For instance, cancer stages and treatment processes involve references to time.
6. Language structures: language structures, such as sentence structure and punctuation, can be indicators of social support. For example, sentences that start with the pronoun “I” are likely to be self-disclosing. And when a writer is asking direct questions, the sentences often end with question marks or use a verb-subject order (eg, “can you”).

To operationalize the linguistic features in the 6 domains, we measured 3 kinds of low-level language features as summarized in Textbox 1.

Textbox 1. Summary of the 3 kinds of basic language features.

1. Linguistic Inquiry and Word Count (LIWC) dictionaries

- Pronoun: I, we, you, she/he, they, impersonal pronoun
- Tense: auxiliary verb, past, present, future
- Emotion: positive emotion, negative emotion (anxiety, anger, sadness)
- Other topics: cognitive mechanism, biological processes, time, religion, death

2. Latent Dirichlet allocation (LDA) topical dictionaries

- Prediagnosis, treatment plan, forum communication, adjusting to diagnosis, financial concerns, lymphedema, diet, family and friends, positive life events, surgery, thoughts and feelings, chemoradiation, family history, emotional reaction, tumor treatment, spiritual, emotional support, routine and schedule, hair loss and appearance, postsurgery problems

3. Syntactic and sentiment features

- Length: sentence count, word count per sentence
- Negation: not
- Part-of-speech: proper nouns, adjectives, cardinal numbers
- Advice pattern: advice verbs, <Please + VERB>, <If + you>, <You + MODAL>
- Question pattern: question marks, any, <VERB + SUBJECT>, indirect questions
- State: be verbs, stative verbs, <SUBJECT_I>, <SUBJECT_I + positive_ADJECTIVE>, <SUBJECT_I + negative_ADJECTIVE>
- Subjectivity: strong subjectivity, weak subjectivity
- Drug: FDA drugs

The first type of language feature was derived from generic dictionaries developed by Pennebaker and colleagues [36] in the Linguistic Inquiry and Word Count (LIWC) program, which measures function words (especially various types of pronouns) and topics with psychological relevance (eg, positive emotion words, negative emotion words, cognition words). Second, we created specialized, cancer-related dictionaries using latent Dirichlet allocation (LDA) topic modeling. Third, we included syntactic features of the messages. Finally, because storytelling is itself a complicated language feature, we included a feature that indicates whether a message reveals a story about the writer.

The LIWC program is a word counting program that calculates the frequency with which words in a text match each of 68 predefined dictionaries representing linguistic dimensions (eg, pronouns, tense), psychological constructs (eg, positive emotion), and personal concerns (eg, leisure, death) [36]. Alpers and colleagues [23] analyzed several hundred posts in an online breast cancer community using a human rater and LIWC and demonstrated a moderate correlation between the ratings assigned by the rater and LIWC scores. Motivated by their work, we included LIWC scores in our machine learning models and considered them as baseline features. LIWC dictionaries were selected based on their a priori relevance to social support concepts. For example, for emotional support, words from the we dictionary (eg, “we,” “us,” “ours”) express feeling of companionship and solidarity, whereas those from the positive emotion (eg, “love”) and religion dictionaries (eg, “pray”) express encouragement. For informational support, words form the impersonal pronoun dictionary (eg, “it”) and the present tense dictionaries are often used to describe objective facts. We included 17 LIWC dictionaries listed in Textbox 1.

The preceding features are generic rather than tailored to the content of cancer-related discussions. Research on analyzing text in support communities suggests that different topics can signal different types of social support interactions [37]. For example, when people use surgery-related terms, such as reconstruction, skin, and surgeon, they are likely to be seeking information, but when people express their fears and other feelings, they are likely to be seeking emotional support. LDA is a statistical generative model that can be used to discover hidden topics in documents and the words associated with each topic [38]. We first trained an LDA model using 30,000 breast cancer messages randomly selected from the entire dataset. The model was set to derive 20 latent topics. For each topic, we chose the 500 words with the highest association with the topic and used them to build a topic dictionary. Two experts familiar with cancer manually assigned a label to each topic (see Table 3). Examples of topics derived from the LDA analysis include emotional reaction (eg, “better,” “lucky,” and “sacred”), diet (“fat,” “weight,” “food,” “exercise,” and “body”), and tumor treatment (eg, “biopsy,” “nodes,” “positive,” and “report”). Textbox 1 lists the 20 topics included in the LDA analysis and Table 3 also shows the words that are most representative of each LDA topic dictionary. Each LDA topical feature calculates the frequency of words in a message matching its corresponding dictionary. When the LDA procedure identified topics that overlapped with preexisting LIWC dictionaries (eg, the LDA family/friends dictionary overlaps the LIWC friends and family dictionaries), we used only one of the dictionaries in the machine learning models for reasons of parsimony.

Table 3. Samples of vocabulary in latent Dirichlet allocation (LDA) topic dictionaries.

LDA topic	Sample vocabulary
Prediagnosis	Told, appointment, wait, back
Treatment plan	Clinical, risk, medicine, therapy
Forum communication	Post, read, help, thread
Adjusting to diagnosis	Understand, trying, experience
Financial concerns	Insurance, plan, company, pay
Lymphedema	Arm, pain, swelling, fluid, area
Diet	Eat, weight, food, exercise, body
Family/friends	Daughter, sister, wife
Positive life events	Love, nice, happy, enjoy, fun
Surgery	Breast, surgeon, mastectomy
Thoughts/feelings	Think, remember, believe
Chemoradiation	Chemo, radiation, treatment
Family history	Mom, children, age, young
Emotional reaction	Better, lucky, scared
Tumor treatment	Biopsy, nodes, positive, report
Spiritual	Love, god, prayer, bless, peace
Emotional support	Hope, hug, glad, sorry, best, luck
Routine/schedule	Today, night, sleep, work
Hair loss/appearance	Hair, wig, grow, head
Postsurgery problems	Pain, blood, tamoxifen, symptom

Sentence count and word count per sentence were features designed to represent the length and complexity of messages. The negation feature is the number of sentences in a message containing negation words or phrases, such as “not,” “shouldn’t,” or “did not.” Because some parts of speech (POS) can signal disease-relevant information or emotion, we counted the number of several specific POS tags. For instance, professional labels can be signaled as proper nouns (eg, “Dr. Smith”), emotional states can be signaled by adjectives (eg, “happy” life), and numbers can be used to describe symptoms or treatments (eg, “10 days since the biopsy”). We applied the Stanford POS tagger [39] to assign POS tags for words and extracted relevant POS features. To identify sentences involving requests and advice, we identified several text patterns or verbs in messages. For instance, <Please+VERB> is a pattern that detects sentences beginning with the word “please” followed by a verb (eg, “Please give”). <You+MODAL> is a pattern that counts the number of sentences that start with a pronoun “you” and are immediately followed by a modal verb expressing possibilities (eg, “should,” “might,” “must”). Furthermore, the advice verbs feature considered the occurrence of verbs such as “make,” “suggest,” and “wish.” The question pattern features were designed to count the number of both direct and indirect question sentences in a message. There were 4 question pattern features, including count of question marks, count of indefinite pronouns (eg, “any,” “anyone,” “anybody”), sentences starting with a modal verb (eg, “Does anyone know...”), and indirect questions (eg, “I am wondering if...”). The state features were designed to capture states that last for a while. These features are potentially related to description of emotional states and storytelling. We counted the number of “be” verbs (eg, “My mom is depressed”), stative verbs (eg, “I love my mom”), sentences starting with the subject “I” (ie, <SUBJECT_I>), and sentences starting with the subject “I” followed by positive or negative adjectives (ie, <SUBJECT_I + positive_ADJECTIVE> and <SUBJECT_I + negative_ADJECTIVE>). Sentiment features described the subjectivity of a text segment. We counted the number of strong subjectivity words (eg, “reject,” “nervous”) and weak subjectivity words (eg, “idea,” “suggest”) for every message using the OpinionFinder subjectivity lexicon [40]. Finally, the number of drug terms in each message was counted, based on an exhaustive list of medicine names collected from the Food and Drug Administration website [41].

The binary story feature indicated whether or not the writer was telling a personal story in the message. A linguistic expert categorized 2200 randomly selected messages as a story or not a story. We then built a machine learning model to predict whether the writer of a message was or was not telling a story. The feature set for this model included all the basic language features described previously. The kappa agreement statistic between human coding and computer prediction was .88, which indicates high agreement [42]. We then applied the story model to detect story messages in the 1.5 million messages collected from the support community. In Table 4, the basic language features are mapped onto the language usage domains.

Table 4. Mapping of language features onto language usage domains.

Language features and usage domains	LIWC	LDA topics	Syntax
Body	Biology processes, death	Prediagnosis, treatment plan, adjusting to diagnosis, lymphedema, diet, surgery, chemo radiation, family history, tumor treatment, hairloss and appearance, postsurgery problems	Drug
Positive/negative emotions and events	Sentiment, cognitive mechanism	Financial concerns, positive life events, thoughts and feelings, emotional reaction, emotional support	State, subjectivity
Social	Pronouns	Forum communication, family and friends	—
Spiritual	Religion	Spiritual	—
Time	Time	Routine and schedule	—
Language structure	Tense	—	Length, negation, part-of-speech, advice pattern, question pattern

Steps 3 and 4: Construction and Performance of Machine Learning Models

Our task was a machine learning regression problem. We built 7 machine learning regression equations to predict in each message the mean of the Turkers' judgments of the amount of emotional or informational support, question asking, or self-disclosure (see the judgment dimensions listed in Table 5). The predictor variables were the dictionaries and other features listed in Textbox 1 and the story feature. We used Weka, a machine learning toolkit, to build the support vector machine regression models (SMOreg) [35]. The 1000 thread starters or their first replies coded by MTurk workers were randomly partitioned into a training set (80%), a development set (10%), and a test set (10%). The training set was used to build the models. The development set was used to evaluate the accuracy of different configurations of the models and variations in the features used. Once the models achieved good performance on the development data, we used the test set to evaluate how well the final regression equations performed. We evaluated the predictions using the Pearson product moment correlation between the human-coded ratings and machine measurements for the 100 messages in the test sample. The agreement between the human-coded ratings and machine measurements was .65 averaged across the 7 dimensions and ranged from .85 for informational support provision to .44 for positive emotional self-disclosure. Table 5 shows the evaluation results for each support-related construct.

Table 5. Accuracy and features for 7 machine learning models.

Machine learning model and top 10 features	Accuracy (Pearson *r*)[a]	Feature weight of SMOreg[b]
Positive emotional self-disclosure	.44	
Positive emotion		0.32
Word count per sentence		0.28
Religion		0.25
<Please + VERB>		–0.21
Sentence count		0.16
<SUBJECT_I + positive_ADJECTIVE>		0.13
Negation		–0.10
We		0.07
Financial concerns		–0.07
Strong subjectivity		0.07
Negative emotional self-disclosure	.59	
Anxiety		1.18
Anger		0.51
<SUBJECT_I>		0.40
Sadness		0.28
<SUBJECT_I + negative_ADJECTIVE>		0.27
Death		0.23
Negation		0.18
Strong subjectivity		0.17
Word count per sentence		0.14
Sentence count		0.14
Positive informational self-disclosure	.45	
Positive emotion		0.31
Religion		0.27
Sadness		–0.25
Sentence count		0.25
Word count per sentence		0.23
<Please + VERB>		–0.20
<SUBJECT_I + positive_ADJECTIVE>		0.16
Routine and schedule		0.13
Biological processes		–0.13
Auxiliary verb		–0.12
Negative informational self-disclosure	.64	
Anxiety		0.42
Sentence count		0.41
Any		0.32
Biological processes		0.28
Tumor treatment		0.26
<SUBJECT_I>		0.26
<SUBJECT_I + positive_ADJECTIVE>		–0.25
Anger		0.24

Machine learning model and top 10 features	Accuracy (Pearson *r*)[a]	Feature weight of SMOreg[b]
I		0.23
Lymphedema		0.21
Question asking	.78	
Sentence count		–0.82
Religion		–0.72
Word count per sentence		–0.64
Positive emotion		–0.59
Question marks		0.52
Any		0.50
Proper nouns		–0.40
<Please + VERB>		0.36
Spiritual		–0.30
Negation		0.27
Emotional support provision	.81	
Sentence count		0.55
Emotional support		0.46
We		0.45
She/He		–0.44
You		0.37
Question marks		–0.33
Strong subjectivity		0.24
Adjusting to diagnosis		0.23
Be verbs		0.23
Positive life events		–0.23
Informational support provision	.85	
Sentence count		1.13
Word count per sentence		0.38
Question marks		–0.33
Spiritual		–0.26
Postsurgery problems		0.22
I		–0.20
<If + you>		0.20
Strong subjectivity		–0.19
Forum communication		–0.17
Tumor treatment		0.16

[a] The accuracy correlation is the Pearson product moment correlation between the average of 10 human judgments and the output of the machine learning model.

[b] The output feature weight of the support vector machine regression model shows the strength of the association between the presence of a feature in a message and human judgments of that message.

Table 5 also presents the 10 most important features associated with each model. Each number in the table represents an output weight of the support vector machine regression model showing the strength of the association between human judgments and machine predictions. For example, the occurrence of religion words from LIWC was a positive indicator of positive emotional self-disclosure (weight=.25) but a negative predictor of asking questions (weight=–.72). In Table 5, features listed in angle brackets indicate patterns (eg, SUBJECT_I indicates a pattern with the pronoun “I” in the subject position in the sentence), whereas other features are names of LIWC or LDA-derived dictionaries.

Given the adequate validity of all these models, we then applied them to measure the amount of emotional and informational support, self-disclosure, and question asking for each of the 1.5 million messages in our dataset (Step 5).

Results

Machine-Coded Versus Human-Coded Data

Using machine-coded data from 58,357 thread starters that received at least 1 reply, we sought to partially replicate the analyses presented in part 1 (Figure 2) which tested the hypotheses that self-disclosure elicits emotional support whereas questions elicited informational support. We did not attempt to replicate the mediation analysis from Figure 3 using machine-coded data because the lower-level language features predicting perceived support needs would be the same as those predicting self-disclosure and question asking, producing problems of common method variance. Table 6 shows the descriptive statistics and correlations among the variables used in the model.

Table 6. Descriptive statistics and correlations among constructs using machine learning (N=58,357 discussion threads).

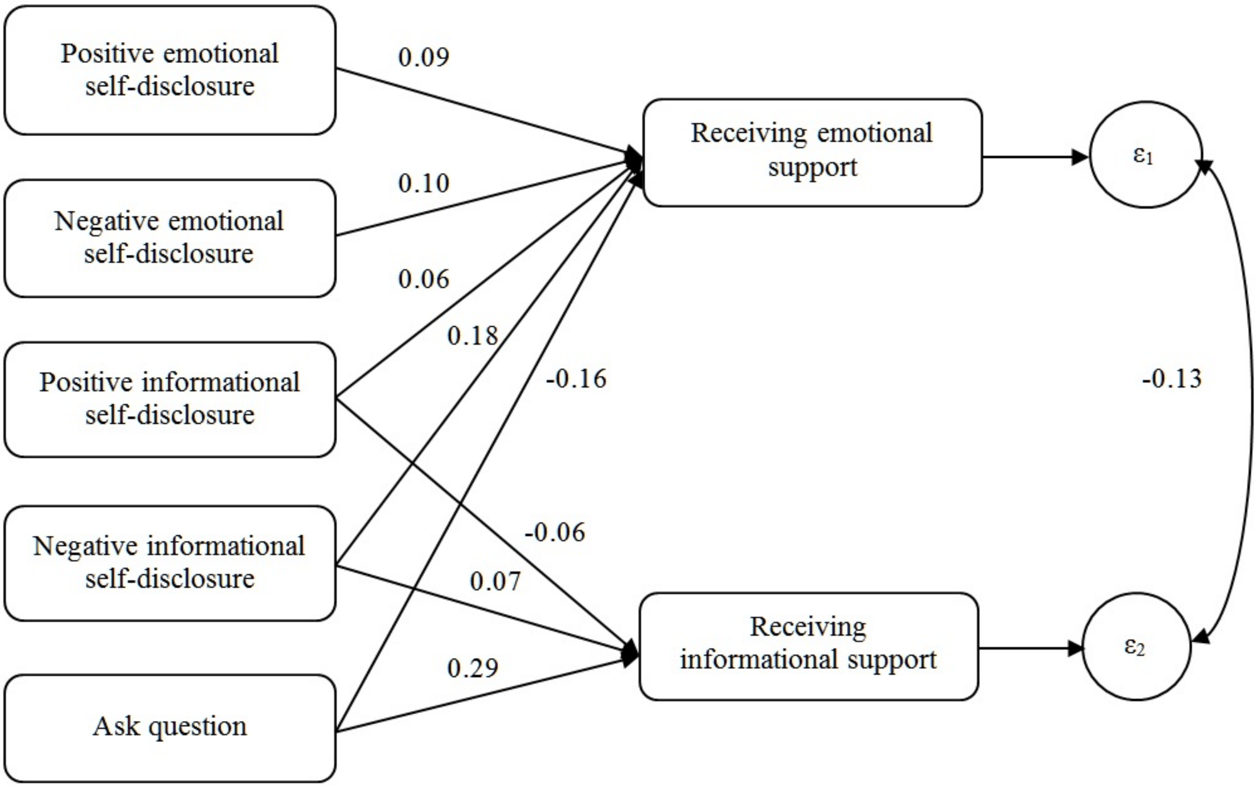
Variable[a]	Mean (SD)	1	2	3	4	5	6
1. Positive emotional disclosure	1.27 (0.26)	1					
2. Negative emotional disclosure	2.29 (1.08)	.19	1				
3. Positive informational disclosure	1.67 (0.41)	.83	.14	1			
4. Negative informational disclosure	3.69 (1.26)	–.05	.67	.04	1		
5. Question	5.26 (1.29)	–.53	.16	–.48	.50	1	
6. Provide emotional support	2.64 (1.06)	.24	.23	.24	.17	–.13	1
7. Provide informational support	2.95 (1.18)	–.22	.08	–.20	.21	.36	–.16

[a] All variables were estimates from machine learning models predicting the amount a message contained the constructs in the first column, rated on a 7-point Likert scale where 1=not at all and 7=very much. The self-disclosure and question measures were based on the thread-starting message, whereas the measures of emotional and informational support were based on the first reply received in the thread.

Figure 4 is the path model showing the results in which the language features in the thread-starting posts were used to predict the amount of emotional and information support provided in the first reply. The results reported are similar to those obtained if we used the average amount of support provided in the first 5 replies. This model fit the data very well according to both the CFI (CFI=0.998) and the RMSEA (RMSEA=0.016).

Most results from this large sample of machine-coded data replicated those from the small sample of human-coded data. The model showed that the more any of the 4 types of self-disclosure occurred in the thread-starting message, the more the first reply contained emotional support (all *P*<.001), although the effect for positive emotional self-disclosure was not significant (*P*=.10) in the model of human-coded data. Specifically, in eliciting emotional support, the effect of negative informational self-disclosure (beta=.18, SE 0.01) was stronger than the effect of either positive emotional self-disclosure (beta=.09, SE 0.01) or negative emotional self-disclosure (beta=.10, SE 0.01), which in turn were stronger than the effect of positive informational self-disclosure (beta=.06, SE 0.01). However, these comparisons of effect strength should be treated with caution given high correlations among some constructs in Table 6 (eg, the correlation between positive emotional and positive informational self-disclosure was .83). As with the human-coded dataset, when the thread starter asked questions, the reply contained less emotional support (beta=–.17, SE 0.01) but more informational support (beta=.29, SE 0.01). Positive informational self-disclosure seemed to depress informational support (beta=–.06, SE<0.00). Negative informational self-disclosure seemed to elicit informational support (beta=.07, SE<0.00), although this association was not significant in the human-coded dataset, presumably because of the small sample size.

Figure 4. Path model showing the analysis of the social support communication process based on machine-coded data. ε1 and ε2 indicate error terms. All P values <.001.



Discussion

Results from both parts 1 and 2 are consistent with the thesis that different conversational strategies lead to receiving different types of social support. People received information and advice when they explicitly requested it, whereas they received emotional support when they disclosed more details about events in their lives and their thoughts and feelings, especially negative ones. Using distinct methods—a small sample of hand-coded data and a large sample of machine-coded data—these parts of the research reached similar conclusions about the relationship between language people used in soliciting support and the kind of support they received. However, conclusions drawn from machine learning models can be more precise because of their larger sample sizes. For example, the sign for the influence of positive emotional self-disclosure on receiving emotional support was the same in the human-coded and machine-coded datasets but only achieved statistical significance in the latter because of the greater statistical power associated with a much larger sample of conversations. These results suggest that machine coding of conversational data can add value beyond what can be learned from human-coded data.

Why is emotional support asked for indirectly, whereas informational support is requested directly? One explanation for the greater use of indirect strategies for soliciting emotional support is that the support may be perceived as less valuable if one needs to ask for it. A major goal of emotional support is to show that the provider cares for the recipient. Simply noticing that the recipient needs support and providing it may itself evidence of caring. Therefore, relatively spontaneous offers of emotional support may be seen as demonstrating more caring than do responses that were explicitly requested by the recipient. In contrast, the value of information or advice is based on the quality of the information offered and not how it was solicited. The quality of this information is not diminished if it is explicitly asked for.

A related explanation is based on the causal attributions people are likely to make if they fail to receive the support they seek. Our data show that people generally receive the type of support they solicited, with moderate correlations between seeking and receiving informational support (r=.37) and between seeking and receiving emotional support (r=.31). However, people sometimes received support that did not match their intent or received no responses at all. If someone seeks informational support from the community and does not receive it, a plausible attribution is that no one in the community who read the request knew the answer. However, because responses to emotional support requests may be seen as generic expressions of concern, a plausible attribution for failing to receive emotional support when it was requested is that no one in the community cared enough to offer an expression of concern. Thus, requesters might hedge their emotional requests to maintain the illusion that they are “care worthy” even if they fail to receive support and perhaps also to maintain the illusion that other community members are “caring” even if they fail to provide support [43].

A final reason for why people ask explicit questions to elicit informational support but ask for emotional support more implicitly is that people seeking informational support may have an easier task defining what they are looking for than do those seeking emotional support. Informational support seekers want answers to specific questions about treatments, medications, symptoms, and problems of daily living. For example, in our corpus, they wanted to know if it is dangerous to own a cat when one has breast cancer, how long depression lasts after a diagnosis of breast cancer, who are the good doctors in a location, or whether to get reconstructive surgery after breast removal. These people know what kind of information or advice they are seeking and hence can ask for it directly. However, people seeking emotional support may be uncertain about exactly what responses they want. When seeking emotional support, people use phrases such as “I am frightened about the side effects of the chemo and I’m worrying about how it will affect me” or “My friend has just been diagnosed with breast cancer, she is not even 32 years old” or “Terrified of what I have. I am 8 months past radiation and I have this weird swelling on the side of my ribs.” Although these people may have strong needs for comfort, reassurance, or social comparison, they may not be able to articulate the form of support that would relieve their distress. As a result, they may disclose their inner turmoil or the negative events in their lives but fail to request specific types of support because they do not know what to ask for. Instead, they leave it to other members of the community to diagnose their needs and respond appropriately.

Part 3: Predicting Commitment to a Support Community From Receiving Social Support

Overview

Although part 2 of the research used machine learning methods with a large data sample to replicate and extend findings from part 1 using a small sample, part 3 of the research shows the value of automated coding by addressing a research question that cannot be practically answered with a small sample—namely, how exposure to support predicts group members’ subsequent participation in the group. Here we briefly summarize a previously published study in which the automated measurement of emotional and informational support described in part 2 was used to predict participants’ length of participation (survival) in their breast cancer support community [25].

Length of participation, an index of commitment to a group, is an important topic for study because it affects the outcomes of individual members as well as the success of the group as a whole. People who stay in an online support community longer are more likely to receive whatever benefits it provides. Moreover, members who remain in the group are resources to fellow members. They share information, provide help, and form social ties with others, and shift over time from being seekers of support to being providers [13,44]. The literature on commitment to communities suggests that both informational and emotional support should increase commitment because participants are likely to consider them important benefits of participation and evidence indicates that received and anticipated benefits increase the likelihood that members will remain in the group [45].

Methods

Overview

We applied survival analysis to test the hypothesis that people who were exposed to more support remained in support communities longer, controlling for the nonsupportive communication they received. Survival analysis is a statistical technique for investigating influences on time-related outcomes, such as whether and when an event occurs. In this study, the event of interest was the time until a member left the community (or conversely, the length of time the person continued participating). The analysis predicted length of participation in support communities from exposure to emotional and information support and other control variables. Because in most online groups the probability of leaving is much higher early in members’ tenure in the group than later on, we used parametric regression survival analysis with time varying independent variables, assuming a Weibull distribution of survival times. The analysis was implemented using the Stata version 13 (StataCorp LP, College Station, TX, USA) streg maximum likelihood estimation for parametric regression survival-time model with multiple records per participant.

Sample

The data we collected did not contain information about which messages people read, but only those they posted. To estimate the amount of support people were exposed to, we assumed that people read all the messages in the threads to which they posted in the week they posted. Thus, the survival analysis was limited to the 30,301 people who posted at least 1 message because without overt behavior it was impossible to estimate the amount of support that they viewed. With this volume of data—approximately 30,000 members exposed to approximately half a million messages—it would be impractical to manually code all the posts. Therefore, our measures of exposure were based on the machine-coded measures of emotional and information support described in part 2.

Measuring Dropout

We considered the timestamp of the first post by each member as the starting date for participating in the breast cancer discussion forums. Moreover, we assumed that a participant left the community if they failed to post again within 12 weeks of their last post. According to these criteria, a user could drop out from the group and rejoin it multiple times. The results reported subsequently were the same if we assumed that people left the group only once. Because people whose last post was within 12 weeks of the end of data collection could still be participating, we treated them as right censored.

Predictor Variables

Post count exposure was the total number of posts to which a user was exposed, assuming the user read all the messages posted in threads during the week the user posted. It is the number of posts in the same week and thread in which the user had posted a message.

Exposure to emotional and informational support for a user were the mean levels of machine-coded emotional and informational support, respectively, averaged over all messages posted in the same week and thread in which the user had posted a message.

Has a profile was a binary measure that indicated whether a user created a profile page (1) or not (0). A total of 31.19% (9,452/30,301) had done so.

Percent thread starters was the percentage of an individual's posts in a week that were thread starters (ie, the number of thread starters a user posted in a week divided by the total number of posts for that user in that week). We included this variable because people who start conversations may be different from those who participate in conversations started by others.

Results

Table 7 and Figure 5 show the results of the survival analysis. Effects are reported in terms of the hazard ratio, which is the effect of unit increase in an explanatory variable on the probability of participants' leaving the community in any particular week. Because all explanatory variables except "has a profile" were standardized, the hazard ratio was the predicted change in the probability of dropout for a unit increase in the predictor (ie, has a profile changing from zero to 1 or the continuous variable increasing by a standard deviation when all the other variables are at their mean levels). A hazard ratio greater than 1 indicates an increased probability of leaving, whereas a ratio less than 1 indicates an increased probability of staying.

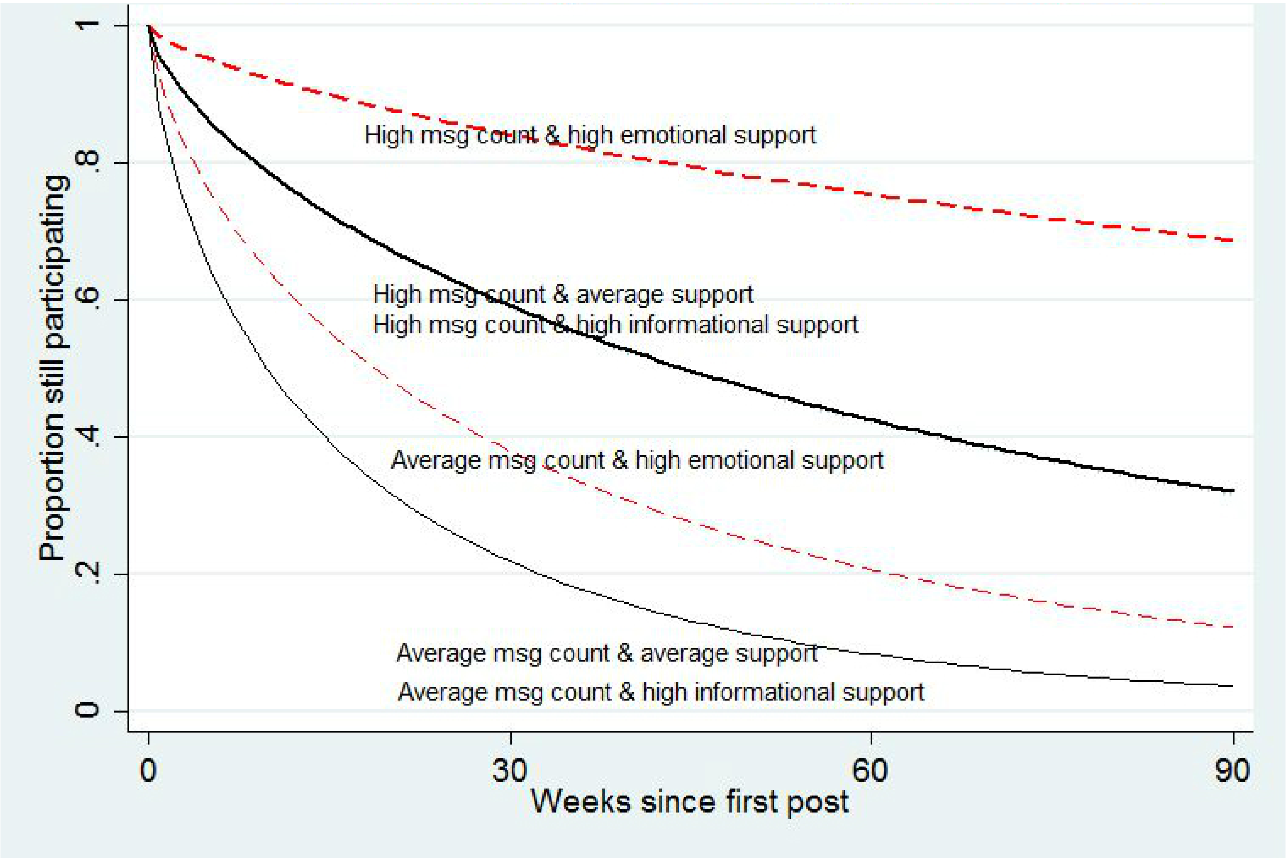
Table 7. Results of the survival analysis.

Predictor variables	Hazard ratio	SE	*P*
Has a profile	0.511	0.010	<.001
% Thread starters	0.853	0.010	<.001
Post count exposure	0.343	0.012	<.001
Emotional support exposure	0.665	0.008	<.001
Informational support exposure	1.048	0.012	<.001
Post count exposure × emotional support exposure	0.493	0.011	<.001
Post count exposure × informational support exposure	0.953	0.020	.02

In interpreting the hazard ratio, the comparison consisted of members with no profiles who were exposed to an average number of messages containing average levels of emotional and informational support. The hazard ratio value of .51 for has a profile meant that members with a profile were 49% more likely to continue to participate than those without a profile (100%–[100%*0.51]). Those who started more threads in a week were 15% more likely to continue participating than those who merely responded to others' posts.

The hazard ratio for post count exposure indicated that survival rates are 66% higher for those who saw a standard deviation more messages than average, when all other variables were at their average levels. Those who were exposed to an average number of messages containing an average of a standard deviation more emotional support were 33% more likely to remain in the group. In contrast, those who were exposed to messages containing an average number of messages with an average of a standard deviation more informational support were 5% more likely to leave the group. The interaction between number of messages and type of support showed that the differential effects of emotional versus informational support were greater when people were exposed to more messages. Those who were exposed to a standard deviation more messages with a standard deviation more emotional support were 50% more likely to remain in the group, whereas those who were exposed to a standard deviation more messages with a standard deviation more informational support were 5% more likely to leave. The effects of being exposed to more messages and to messages with more emotional support were substantively powerful. For example, based on the survival analysis in Table 7, after 90 days only 3% of those exposed to an average number of messages with an average amount of emotional support would be expected to be active in the groups. In contrast, over 10 times as many (35%) would be expected to be active if they were exposed to a standard deviation more messages at average support levels and over 20 times more (67%) would be expected to be active if they were exposed to a standard deviation more messages containing a standard deviation more than average emotional support.

Figure 5. Survival curves for members exposed to different numbers of posts and type of social support. Note: although receiving more informational support was reliably associated with lower longevity on the site, the effect was small and the lines representing high informational support cannot be visually distinguished from the lines representing average informational support.



Discussion

There are several plausible explanations for why emotional support was associated with remaining in the group whereas informational support was associated with leaving. It may be that many informational needs are short-term. As a result, people who have informational needs and receive information from others have these immediate needs met and have little reason to return, just as one might not continue perusing a dictionary after looking up a definition. On the other hand, the need for emotional support may be longer term and require multiple interactions to be fulfilled. Another possibility is that people who seek emotional support are themselves more social and predisposed to appreciate the encouragement and social comparison the community offers, whereas seekers of information are not. Finally, factual information exchanged in unmoderated health support communities may lack the accuracy, credibility, and usefulness of information from vetted sources, such as physicians or websites run by the American Cancer Society. For this reason, people exposed to informational support may leave because of their negative evaluation of the information they receive. In contrast, emotional support obtained in a community may be perceived as more valuable because the social comparisons, empathy, and encouragement from similar others on the site cannot be duplicated outside the group.

Summary and Conclusion

The exchange of social support in online support communities is a dynamic communication process in which people actively seek out and then receive support. This study shows that different conversational moves are associated with receiving emotional and informational support in a cancer support community. When people self-disclose about their experiences, especially negative ones, and when they report on their negative thoughts and feelings, others provide them with emotional support. In contrast, when people ask questions and talk about negative but not positive events in their lives, others provide them with informational support. In addition, the support people receive is associated with their continued participation in a support community. They are more likely to continue participating after being exposed to more messages and to messages with higher emotional support, but they are more likely to leave after being exposed to messages with higher informational support.

In addition to these substantive results, our research demonstrated the value of automated coding of online conversations. Most studies of communication in support communities are based on hand coding relatively small samples of conversations. Using machine learning techniques in part 2, we were able to largely replicate findings based on human coding and also to discover relationships that underpowered human-coding studies did not find. In addition, part 3 used automated coding to conduct analyses that would be infeasible with human coding; namely, examining how exposure to more than half a million messages over a 10-year period is associated with over 30,000 participants' commitment to online support communities.

We also provided an overview of how to use modern machine learning and language analysis techniques to better understand interactions in online health support communities and other online communities. Although previous researchers have shown that it is feasible to automate some text analysis of conversations in online support communities and to produce machine coding results that partially mirror human-coding results, most failed to provide enough details about their methods to allow a skeptical researcher to challenge assumptions they made or to replicate their work. To rectify this shortcoming in the literature, we provided substantial information about machine learning models designed to automatically identify self-disclosure, question asking, and emotional and informational support in breast cancer discussion forums. The performance of our models for predicting social support concepts shows that it is feasible to use computer programs to automatically analyze conversations in online support communities. In particular, we believe that the feature set we used can be fruitfully applied to build predictive models for social support using other health support datasets. LIWC and the linguistic features can be directly applied because they are generic in the sense that they are not tailored to any specific domain. The main work researchers would need to do when applying these methods to another health domain is to recreate the LDA topical dictionaries customized to the data of interest. Many tools now exist for conducting an LDA analysis of a corpus [46,47], and this step can be applied to a new corpus with little effort.

Limitations and Future Work

Some limitations of the current research should be noted. Regarding the automated coding, accuracy for some constructs, especially the self-disclosure measures, was lower than desirable, thereby attenuating effect sizes. Although it is possible to adjust effect size estimates for attenuation [48], future efforts should be made to improve the measurements, perhaps by adding additional language features or features reflecting the posters' past behavior. In addition, the ground truth labels for developing the machine learning models were based on annotations provided by MTurk workers who were not the people requesting or receiving support. A natural next step is to validate the labels provided by the MTurk workers with labels provided by the people actually providing or receiving support.

Furthermore, although our current analysis was based on a large corpus of data from 66 forums, we only examined a single disease (breast cancer) in a single online health support site. Other online health support communities might produce different patterns. For example, the positive effect of self-disclosure on emotional support provision in a prostate cancer support community might be different because men are less likely to talk about their thoughts and feelings and, when they do self-disclose, their behavior might be more notable. Further research studying other online support communities can help us better understand and confirm our findings.

Finally, although our findings suggest that people use different language strategies to get informational and emotional social support and that exposure to informational and emotional social support has different effects on commitment, our analyses are correlational. We know that solicitation attempts precede the provision of support and that exposure to support precedes the decision to leave or stay in the group, but we have not demonstrated either that the messages we labeled as support elicitations are actually requesting support, that the messages we labeled as providing support actually do so, or that language in thread-starting messages or replies actually causes others to act differently. Laboratory or field experiments that manipulate support language are needed to test whether the correlational relationships we observed represent causal relationships.

Although the current research used automated coding of support language to better understand the conversational dynamics in health support communities, these same techniques could be used to improve the way these communities function. Using filtering techniques common in online social networking sites, such as Facebook, it would be possible to increase the likelihood that support-seeking messages are seen by those people who are able to provide it. For example, using the automatic classification procedures we have described, it would be possible to automatically identify messages that are seeking a certain type of support but that do not receive it during the 24-hour window when over 80% of messages receive replies. These messages could then be sent by email to others in the community who have been recently active and have responded to similar messages in the past.

Acknowledgments

This research was supported by NSF grant IIS-0968485. We want to thank Dong Nguyen who helped provide the data for this project. Carolyn Rosé provided valuable feedback on our research and writing.

Conflicts of Interest

None declared.

References

1. Chou WYS, 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]

2. Chou WYS, Liu B, Post S, Hesse B. Health-related Internet use among cancer survivors: data from the Health Information National Trends Survey, 2003-2008. J Cancer Surviv 2011 Sep;5(3):263-270. [doi: 10.1007/s11764-011-0179-5] [Medline: 21505861]
3. Rains SA, Young V. A meta-analysis of research on formal computer-mediated support groups: examining group characteristics and health outcomes. Human Communication Research 2009;35(3):309-336. [doi: 10.1111/j.1468-2958.2009.01353.x]
4. Lieberman M, 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]
5. 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]
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. British Medical Journal 2004;328(7449):1166-1171. [doi: 10.1136/bmj.328.7449.1166]
7. Lakey B, Lutz C. Social support and preventive and therapeutic interventions. In: Handbook of Social Support and the Family. New York: Springer US; 1996:435-465.
8. Ridings CM, Gefen D. Virtual community attraction: why people hang out online. Journal of Computer-Mediated Communication 2004 Nov;10(1):00 [FREE Full text] [doi: 10.1111/j.1083-6101.2004.tb00229.x]
9. Dunkel-Schetter C, Feinstein LG, Taylor SE, Falke RL. Patterns of coping with cancer. Health Psychol 1992;11(2):79-87. [Medline: 1582383]
10. Dunkel-Schetter C. Social support and cancer: findings based on patient interviews and their implications. Journal of Social Issues 1984;40(4):77-98. [doi: 10.1111/j.1540-4560.1984.tb01108.x]
11. Helgeson VS, Cohen S, Schulz R, Yasko J. Group support interventions for women with breast cancer: who benefits from what? Health Psychol 2000 Mar;19(2):107-114. [Medline: 10762094]
12. Jacobs C, Ross RD, Walker IM, Stockdale FE. Behavior of cancer patients: a randomized study of the effects of education and peer support groups. Am J Clin Oncol 1983 Jun;6(3):347-353. [Medline: 6342360]
13. Helgeson VS, Cohen S, Schulz R, Yasko J. Long-term effects of educational and peer discussion group interventions on adjustment to breast cancer. Health Psychol 2001 Sep;20(5):387-392. [Medline: 11570653]
14. Pearlin L, McCall M. Occupational stress and marital support. In: Eckenrode J, Gore S, editors. Stress Between Work and Family. New York: Springer US; 1990:36-60.
15. Barbee AP, Cunningham MR, Winstead BA, Derlega VJ, Gulley MR, Yankeelov PA, et al. Effects of gender role expectations on the social support process. Journal of Social Issues 1993;49(3):175-190. [doi: 10.1111/j.1540-4560.1993.tb01175.x]
16. Bambina A. Online Social Support: The Interplay of Social Networks and Computer-Mediated Communication. New York: Cambria Press; 2007.
17. Burke M, Joyce E, Kim T, Anand V, Kraut R. Introductions and requests: rhetorical strategies that elicit response in online communities. In: Communities and Technologies. London, UK: Springer London; 2007:21-39.
18. Rodgers S, Chen Q. Internet community group participation: psychosocial benefits for women with breast cancer. Journal of Computer-Mediated Communication 2005;10(4):00. [doi: 10.1111/j.1083-6101.2005.tb00268.x]
19. Cutrona CE, Russell DW. Type of social support and specific stress: toward a theory of optimal matching. In: Social Support: An Interactional View. New York: Wiley-Interscience; 1990:319-366.
20. Hardey M. 'The story of my illness': personal accounts of illness on the Internet. Health 2002 Jan 01;6(1):31-46. [doi: 10.1177/136345930200600103]
21. Meier A, Lyons EJ, Frydman G, Forlenza M, Rimer BK. How cancer survivors provide support on cancer-related Internet mailing lists. J Med Internet Res 2007;9(2):e12 [FREE Full text] [doi: 10.2196/jmir.9.2.e12] [Medline: 17513283]
22. 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]
23. Alpers GW, Winzelberg AJ, Classen C, Roberts H, Dev P, Koopman C, et al. Evaluation of computerized text analysis in an Internet breast cancer support group. Computers in Human Behavior 2005 Mar;21(2):361-376. [doi: 10.1016/j.chb.2004.02.008]
24. 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]
25. Wang YC, Kraut R, Levine JM. To stay or leave?: The relationship of emotional and informational support to commitment in online health support groups. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. 2012 Presented at: ACM 2012 conference on Computer Supported Cooperative Work; Feb 11-15, 2012; Seattle, WA p. 833-842.

26. Sebastiani F. Machine learning in automated text categorization. ACM Comput Surv 2002;34(1):1-47. [doi: 10.1145/505282.505283]
27. Chartrand TL, van Baaren R. Human mimicry. Adv Exp Soc Psychol 2009;41:219-274. [doi: 10.1016/S0065-2601(08)00405-X]
28. Niederhoffer KG, Pennebaker JW. Linguistic style matching in social interaction. Journal of Language and Social Psychology 2002 Dec 01;21(4):337-360. [doi: 10.1177/026192702237953]
29. Holler J, Wilkin K. Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. J Nonverbal Behav 2011 Jan 23;35(2):133-153. [doi: 10.1007/s10919-011-0105-6]
30. Snow R, O'Connor B, Jurafsky D, Ng AY. Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2008 Presented at: Conference on Empirical Methods in Natural Language Processing; October 25-27, 2008; Honolulu, HI p. 254-263.
31. Bruckman A. Teaching students to study online communities ethically. Journal of Information Ethics 2006 Nov 1;15(2):82-98. [doi: 10.3172/JIE.15.2.82]
32. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods of Psychological Research Online 2003;8(2):23-74.
33. Arguello J, Butler BS, Joyce E, Kraut R, Ling KS, Rosé C, et al. Talk to me: foundations for successful individual-group interactions in online communities. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2006 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 22-27, 2006; Montréal, QC. [doi: 10.1145/1124772.1124916]
34. Burke M, Kraut R, Joyce E. Membership claims and requests: conversation-level newcomer socialization strategies in online groups. Small Group Research 2009 Dec 11;41(1):4-40. [doi: 10.1177/1046496409351936]
35. Witten IH, Frank E, Hall MA. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems). San Francisco, CA: Morgan Kaufmann; 2011.
36. Pennebaker JW, Booth RJ, Francis ME. Linguistic Inquiry and Word Count: LIWC2007. Austin, TX: LIWC.net; 2007.
37. Buis LR. Emotional and informational support messages in an online hospice support community. Comput Inform Nurs 2008;26(6):358-367. [doi: 10.1097/01.NCN.0000336461.94939.97] [Medline: 19047885]
38. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022.
39. Toutanova K, Manning CD. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In: Proceedings of the 2000 Joint SIGDAT conference on empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics - Volume 13. 2000 Presented at: Joint SIGDAT conference on empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics; October 7-8, 2000; Hong Kong p. 63-70.
40. Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2005 Presented at: Conference on Empirical Methods in Natural Language Processing; October 6-8, 2005; Vancouver, BC p. 347-354. [doi: 10.3115/1220575.1220619]
41. US Food and Drug Adminstration. 2013. Drugs URL: http://www.fda.gov/Drugs/ [accessed 2014-05-22] [WebCite Cache ID 6PlnaGFNM]
42. Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960 Apr 01;20(1):37-46. [doi: 10.1177/001316446002000104]
43. Goffman E. On face-work; an analysis of ritual elements in social interaction. Psychiatry 1955 Aug;18(3):213-231. [Medline: 13254953]
44. Preece J, Shneiderman B. The reader-to-leader framework: motivating technology-mediated social participation. AIS Transactions on Human-Computer Interaction 2009;1(1):13-32.
45. Moreland RL, Levine JM. Socialization in organizations and work groups. In: Turner ME, editor. Groups at Work: Theory and Research. Mahwah, NJ: Psychology Press; 2001:69-112.
46. MALLET. 2014. MAchine Learning for LanguagE Toolkit URL: http://mallet.cs.umass.edu/index.php [accessed 2014-05-22] [WebCite Cache ID 6PlmNBAzg]
47. The Stanford Natural Language Processing Group. 2014. Stanford Topic Modeling Toolbox URL: http://nlp.stanford.edu/downloads/tmt/tmt-0.4/ [accessed 2014-05-22] [WebCite Cache ID 6PlmvW9F2]
48. Armstrong B. Measurement error in the generalised linear model. Communications in Statistics - Simulation and Computation 2007 Jun 27;14(3):529-544. [doi: 10.1080/03610918508812457]

Abbreviations

CFI: comparative fit index
HIT: Human Intelligence Tasks
HR: hazard ratio
LDA: latent Dirichlet allocation

LIWC: Linguistic Inquiry and Word Count
POS: parts of speech
RMSEA: root mean squared error of approximation

Edited by G Eysenbach; submitted 28.05.14; peer-reviewed by J Huh, A Bambina; comments to author 18.06.14; revised version received 05.09.14; accepted 07.10.14; published 20.04.15.

Please cite as:
Wang YC, Kraut RE, Levine JM
Eliciting and Receiving Online Support: Using Computer-Aided Content Analysis to Examine the Dynamics of Online Social Support
J Med Internet Res 2015;17(4):e99
URL: http://www.jmir.org/2015/4/e99/
doi:10.2196/jmir.3558
PMID:25896033

Original Paper

Using a 3D Virtual Supermarket to Measure Food Purchase Behavior: A Validation Study

Wilma Elzeline Waterlander[1], BSc, MSc, PhD; Yannan Jiang[1], BSc, MSc, PhD; Ingrid Hendrika Margaretha Steenhuis[2], MSc, PhD; Cliona Ni Mhurchu[1], BSc (Hons), PhD

[1]National Institute for Health Innovation, School of Population Health, University of Auckland, Auckland, New Zealand

[2]Department of Health Sciences and the EMGO Institute for Health and Care Research, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Wilma Elzeline Waterlander, BSc, MSc, PhD
National Institute for Health Innovation
School of Population Health
University of Auckland
Tamaki Campus
261 Morrin Road
Auckland, 1072
New Zealand
Phone: 64 09 9234612
Fax: 64 93731710
Email: w.waterlander@auckland.ac.nz

Abstract

Background: There is increasing recognition that supermarkets are an important environment for health-promoting interventions such as fiscal food policies or front-of-pack nutrition labeling. However, due to the complexities of undertaking such research in the real world, well-designed randomized controlled trials on these kinds of interventions are lacking. The Virtual Supermarket is a 3-dimensional computerized research environment designed to enable experimental studies in a supermarket setting without the complexity or costs normally associated with undertaking such research.

Objective: The primary objective was to validate the Virtual Supermarket by comparing virtual and real-life food purchasing behavior. A secondary objective was to obtain participant feedback on perceived sense of “presence” (the subjective experience of being in one place or environment even if physically located in another) in the Virtual Supermarket.

Methods: Eligible main household shoppers (New Zealand adults aged ≥18 years) were asked to conduct 3 shopping occasions in the Virtual Supermarket over 3 consecutive weeks, complete the validated Presence Questionnaire Items Stems, and collect their real supermarket grocery till receipts for that same period. Proportional expenditure (NZ$) and the proportion of products purchased over 18 major food groups were compared between the virtual and real supermarkets. Data were analyzed using repeated measures mixed models.

Results: A total of 123 participants consented to take part in the study. In total, 69.9% (86/123) completed 1 shop in the Virtual Supermarket, 64.2% (79/123) completed 2 shops, 60.2% (74/123) completed 3 shops, and 48.8% (60/123) returned their real supermarket till receipts. The 4 food groups with the highest relative expenditures were the same for the virtual and real supermarkets: fresh fruit and vegetables (virtual estimate: 14.3%; real: 17.4%), bread and bakery (virtual: 10.0%; real: 8.2%), dairy (virtual: 19.1%; real: 12.6%), and meat and fish (virtual: 16.5%; real: 16.8%). Significant differences in proportional expenditures were observed for 6 food groups, with largest differences (virtual – real) for dairy (in expenditure 6.5%, $P<.001$; in items 2.2%, $P=.04$) and fresh fruit and vegetables (in expenditure: –3.1%, $P=.04$; in items: 5.9%, $P=.002$). There was no trend of overspending in the Virtual Supermarket and participants experienced a medium-to-high presence (88%, 73/83 scored medium; 8%, 7/83 scored high).

Conclusions: Shopping patterns in the Virtual Supermarket were comparable to those in real life. Overall, the Virtual Supermarket is a valid tool to measure food purchasing behavior. Nevertheless, it is important to improve the functionality of some food categories, in particular fruit and vegetables and dairy. The results of this validation will assist in making further improvements to the software and with optimization of the internal and external validity of this innovative methodology.

(J Med Internet Res 2015;17(4):e107) doi:10.2196/jmir.3774

KEYWORDS

virtual reality; user-computer interface; software validation; nutrition policy; food; behavior; public health

Introduction

Supportive environments are essential for people to make healthier food choices and there is a growing call to implement structural interventions such as fiscal policies and front-of-pack labeling to create a healthier food environment [1-3]. Supermarkets form a key setting for health-promoting interventions [4,5] because they have a central position within the food chain [6,7] and they are where people buy most of their food (87% of households in New Zealand buy foods from supermarkets weekly or more often [8]). Although structural interventions in the retail setting have clear potential, high-quality evidence on their effectiveness is limited [9,10]. Evidence from randomized controlled trials (RCTs) is particularly lacking. Although examples of high-quality supermarket trials exist [10-12], experimental studies have mostly been conducted in settings with a limited number of food choices and most lacked objective outcome measures [13].

The main reason for the dearth of high-quality supermarket trials is that they are complex and costly to conduct. This is especially true for RCTs (eg, it would be problematic to raise prices on soft drinks [simulating a tax] in some supermarkets but not in others). Furthermore, opposition and extensive lobbying from the food industry against several proposed intervention strategies is another important reason for the lack of experimental studies [14].

In order to find a solution to the complexities of undertaking supermarket intervention research, we developed a research tool that can be used to study the effects of interventions in a virtual reality setting: the Virtual Supermarket (see Figure 1). Virtual environments are computer-generated models, which participants can experience and interact with intuitively in real time [15]. Gaming technology is used to simulate a real supermarket shopping experience where study participants purchase virtual food items. Photographs of real foods are used to compose virtual food products and prices are displayed on virtual shelf labels. Currently, 2 versions of this Virtual Supermarket exist: the original Dutch version [16] and a New Zealand version (developed using a similar methodology) that is the subject of this study. Other examples of virtual supermarkets can be found in the scientific literature; however, to our best knowledge, these have different functionalities [17], are developed for other purposes (eg, rehabilitation) [18], or have not published study results yet [19].

To obtain valid study results, it is important that the Virtual Supermarket simulates a real supermarket as closely as possible. Here, one of the most important traits is the level of presence people experience [15]. *Presence* is defined as “a psychological state of ‘being there’ mediated by an environment that engages our senses, captures our attention, and fosters our active involvement” [20] and reflects “the subjective experience of being in one place or environment, even if one is physically located in another” [21]. Factors that are hypothesized to contribute to a sense of presence are (1) involvement, (2) sensory fidelity, (3) adaptation/immersion, and (4) interface quality factors. Second, a key question is whether peoples’ virtual behavior reflects their real-life behavior accurately. Studies using the Dutch Virtual Supermarket have already shown some promising results [22,23]. However, to date, validation of the software has been solely based on self-report (questionnaire) data. Therefore, the aim of this study was to validate the NZ Virtual Supermarket by comparing virtual and real-life food purchasing behavior using objective outcome measures and also to evaluate the level of presence that people experience. Results will inform further development of the software and guide the development of future experiments using the Virtual Supermarket.

Figure 1. Screenshots of the New Zealand Virtual Supermarket.

Methods

The New Zealand Virtual Supermarket

The Virtual Supermarket was developed in Unity (a game development system) and represents a 3-dimensional (3D) computer simulation of a real shopping experience (Figure 1). More details about the software can be found elsewhere [16,24]. The Dutch Virtual Supermarket is available online [25] (the NZ version is currently undergoing modification).

The NZ Virtual Supermarket was designed using the Auckland branch of a market leader as a model. NZ food prices and product range were obtained from a 2011 field survey of packaged foods available for sale in Auckland supermarkets [26] and online supermarkets. Foods were categorized into 18 food groups and further divided into 91 categories, 141 subcategories, and 31 sub-subcategories (eg, dairy—cheese—hard cheese—high salt cheese). The Virtual Supermarket includes a representative number of products in each food category, with at least 1 product in each subcategory, adding up to a total of 1445 unique products (Table 1) (an average NZ supermarket offers approximately 9000 nutritionally unique food products). Within each category, we selected the top-selling products using the Australian Grocery Guide annual report 2010 sales data [27]. In addition, we ensured sufficient variety of special food items such as organic, diet, or gluten-free varieties and included home or discount brands where possible.

In the Virtual Supermarket, participants can navigate around the store with a shopping trolley using their computer cursor keys and select (or remove) groceries with a mouse click. There is an option to view a list of selected groceries, including the price and total amount of money spent. Cash registers are located at the end of the supermarket where participants can check out and complete the study. The software underwent some testing by our study team prior to this validation study, but no formal usability testing was performed.

Table 1. Food groups and food categories within the New Zealand Virtual Supermarket.

Food group and category	Products, n
Alcoholic beverages	
Wine (including cask wine, champagne, dessert wine, red, rose wine, sparkling, special releases, white wine, other liquor)	57
Beer and cider	15
Baby care	
Baby food	13
Baby formula	4
Bread and bakery	
Rolls, bagels, and flat breads	21
Fresh bread (in-store bakery)	4
Sweet treats and cakes	28
Biscuits	73
Bread (packaged)	43
Cakes, muffins, and pastry	26
Baking and cooking	
Herbs and spices	40
Baking additives	12
Bread crumbs and coating	4
Beverages (including coffee, plain; cordial bases; electrolyte drink; energy drink; fruit and vegetable juices; hot drink mixes; ice tea drinks; soft and flavored drinks; tea, plain; water)	115
Cereal and cereal products	
Bran and breakfast cereals	27
Cereal bars	23
Flour	4
Grains	4
Noodles	5
Pasta	31
Rice	15
Sugar	4
Convenience foods	
Pizza	5
Ready meals	23
Ready snacks	10
Soup	18
Dairy	
Cheese	62
Cream	5
Desserts	7
Ice cream	20
Milk	22
Yogurt/ yogurt drinks	20
Deli and chilled foods	
Fresh pasta, sauces, soups, and pesto	4

Food group and category	Products, n
Antipasto and continental meats	12
Fresh meals and sandwiches	3
Chilled salads	3
Sliced meats	9
Edible oils and oil emulsions	
Butter and margarine	11
Cooking oil/spray	9
Eggs	
Free range	2
Cage eggs	2
Barn laid	1
Fish and fish products	
Canned fish	19
Frozen fish	14
Chilled fish	5
Fish with pastry	1
Fresh fruit and vegetables	
Fruit	27
Fresh herbs	7
Green vegetables	7
Prepared salads	5
Other vegetables	22
Fruit and vegetables other	
Fruit (including dried, dried fruit and nut mixes, frozen fruit, fruit bars, fruit in juice/syrup, other)	40
Vegetables (including baked beans, corn [can], legumes [can], tomatoes [can], canned [other], dried legumes, frozen potato products, frozen vegetables)	61
Jam and spreads	11
Nuts and seeds	22
Pickled vegetables	11
Fresh meat and seafood, meat products	
Fresh fish	5
Shell fish	3
Smoked fish	1
Beef and veal (unprocessed)	5
Lamb (unprocessed)	4
Pork (unprocessed)	4
Poultry (unprocessed)	2
Sausages and small good	5
Meat alternatives	6
Processed meat	63
Other miscellaneous	
Other desserts (including frozen juice and ice blocks, fruit crumbles, sponges/pudding, fruit pies, jelly, pavlova/meringues, sorbets, other desserts)	21

Food group and category	Products, n
Sauces, spreads and seasonings	
Mayonnaise/dressings	15
Meal accompaniments	1
Sauces/seasonings	95
Spreads	23
Other savory	1
Pate	6
Peanut butter and sweet spreads	9
Relished and pickles	12
Yeast extract	2
Snack foods	
Crisps and snacks	32
Sweet snacks	54
Chewing gum	7
Total	1445

Study Design

Study participants were asked to shop on 3 occasions in the Virtual Supermarket over 3 consecutive weeks and to collect their real supermarket grocery till receipts for that same period. Virtual and real shops were conducted in the same weeks. Three shopping occasions were included to provide a representative picture of usual food purchases and to test learning effects. Participants were instructed to conduct their weekly virtual shop prior to their real-life shopping trips. This order was chosen because it is unlikely that people adapt their real-life purchases to their choices made in the Virtual Supermarket; however, when real shops are conducted first, it is possible that people could use their real till receipts as a basis for their food purchases in the Virtual Supermarket. The study design and procedures were approved by the University of Auckland Human Participants Ethics Committee (Nov 7, 2012; ID8691).

Participants

Participants were recruited from the general NZ population using advertisements placed in supermarkets, a research participant recruitment website (researchstudies.co.nz), local newspapers, mailing lists and via Māori (indigenous New Zealanders) and Pacific networks within the University. Inclusion criteria were adult (≥18 years), main household shopper, able to communicate in English, access to a computer and Internet, and have an email address. Exclusion criteria included another person in the household already participating in the study.

Study Procedure

Participants completed the study at home or on any computer with an Internet connection. Following registration and consent, an email was sent with study instructions and a URL (link to Web address) to download the Virtual Supermarket. In addition to a hard copy illustrated manual, we recorded an instructional video on the Virtual Supermarket. For all 3 virtual shopping occasions, participants were instructed to purchase food for their household for the coming week as they planned to do in real life. It was emphasized that they should buy *all* food groceries they would need for their household in the coming week, including food they were planning to buy for a special occasion (eg, for a birthday). Likewise, participants were instructed not to buy foods in the Virtual Supermarket if they were not planning to buy these in real life (ie, because they had a sufficient stock). The shopping budget in the Virtual Supermarket was set to the same level as participants' real-life household food shopping budget for that week. After completion of each virtual shop, participants were asked to collect their real supermarket till receipts for that week (a reminder message appeared on their computer screen). Participants were instructed to collect all real-life food shop till receipts, including from supermarkets, speciality stores, convenience stores, and markets, but not from fast food, restaurants, or takeaways. Along with the till receipts, participants completed a questionnaire about their shopping habits that week (eg, "In the past week, did you make any special food purchases that you would not have normally; for example, because there was a sale/promotion or because of a special occasion?") and asking how many till receipts were missing. Participants received email and short message service (SMS) text message reminders throughout the study to remind them to complete their virtual shops and collect their till receipts. Participants who completed all 3 virtual shops and returned their grocery till receipts received a NZ $20 supermarket voucher.

Outcome Measures

Primary Outcome

The primary outcome of this study was the proportional expenditures (NZ$) on each food group (18 food groups, see Table 1) for both the Virtual Supermarket and real supermarket (eg, percentage of total expenditures on vegetables). We measured proportional expenditure instead of absolute expenditures because we were interested in shopping patterns (eg, did participants buy similar quantities of milk and bread) which are better reflected by proportional expenditures (absolute supermarket prices tend to vary substantially over time). Secondly, we also looked at the number of products purchased in each food category and compared the Virtual Supermarket and real supermarket proportions (eg, percentage of all purchased products that are vegetables). Product purchases in the Virtual Supermarket were measured by till receipts automatically generated at the end of each shop. Real-life food purchases were measured by grocery till receipts.

Secondary Outcomes

Secondary outcomes included self-reported level of presence measured by the Presence Questionnaire Items Stems (version 4.0) [20]. This questionnaire includes a total of 29 items that are scored on a 7-point Likert scale. Examples of questions include “How responsive was the Virtual Supermarket to actions that you initiated (or performed)?” and “How much did your experiences in the Virtual Supermarket seem consistent with your real world experiences?” In addition, we measured potential effect modifiers and background variables including:

1. Capability or tendency to be involved or immersed. People can experience different levels of involvement both in virtual tasks and in common activities (focus). This concept was measured with the items that relate to the factor “focus” (7 questions) from the Immersive Tendency Questionnaire Item Stems [21] (ie, “how mentally alert do you feel at the present time?”)
2. Participant characteristics including age, gender, household size, ethnicity, responsibility for household grocery shopping, highest educational qualification, work status, household income, supermarket shopping habits, computer use, and Internet access.

Statistical Analyses

Sample Size

We aimed to recruit 96 participants. This sample size was estimated to provide 85% power at 5% level of significance (2-sided) to detect a minimal absolute difference of 1% in percentage of food products or expenditure on each food category between weekly virtual and real supermarket purchases. This assumed a standard deviation of 3% based on data from previous research [23] and an estimated 20% loss to follow-up rate. Because this was a validation study, the sample size was not adjusted for multiple comparisons.

Comparing Virtual and Real Supermarkets

Data from the virtual and real-life grocery till receipts were entered into SPSS version 20 (IBM Corp, Armonk, NY, USA) and coded within the matching 18 food groups. Data were then imported into SAS version 9.2 (SAS Institute Inc, Cary, NC, USA) for further analysis. For each weekly shopping visit, proportional expenditure (NZ$) and the proportion of products purchased were calculated for each food group. First, alignment of the virtual and the real purchases was described using line graphs displaying mean purchases for all food groups for all participants with a least 1 valid virtual shop (raw data). Second, the resemblance between the virtual and real supermarket was tested using repeated measures mixed models. Both supermarket type and shopping visit were fitted as fixed effects. Their interaction was tested and dropped if it was not statistically significant. A random effect at participant level was also fitted to account for correlation between repeated visits on the same participant, and attrition over time assuming the data were missing at random. Model-adjusted means were estimated for the virtual and real supermarkets, and mean differences were tested. Analyses were conducted on the total sample (ie, including all participants who completed at least 1 virtual shop, N=86), followed by complete-case analysis. This second phase included participants who completed all 3 virtual shops and returned grocery till receipts for all 3 weeks. Also, the till receipts had to be valid and we excluded those who reported having failed to return the majority of their receipts (n=8). To determine missing till receipts, we asked about the number of shops participants conducted during the week, whether these were small or bigger shopping episodes, and how many till receipts they collected. If participants reported having conducted a large shop, this till receipt had to be available for the participant’s data to be included in analysis. If participants did multiple smaller shops, at least 60% of receipts had to be available. This resulted in a total of 52 participants (60%, 52/86 for the complete-case analysis).

Questionnaire Data

After each virtual shopping occasion, participants were asked to complete the presence and focus questionnaires. Because there was no difference in these scores over time, the mean total presence and focus score over all completed shops was derived for each participant (typical value). These mean values were divided into 3 groups (low/medium/high) for descriptive analysis. Linear mixed models were used to test the predicted effect of focus on presence (continuous score). Sensitivity analysis was conducted to test whether participants with a higher presence score performed better in the Virtual Supermarket within the complete cases by selecting a subset of participants with a high sense of presence using the median score (median 100) as a cut-off point.

Results

Participant Characteristics

A total of 309 participants registered for this study via phone or email. The first 135 participants were phoned to check their eligibility; 4 were not eligible because they were unavailable during the study period. The remainder were sent the baseline questionnaire and consent forms, with 123 participants enrolled in the study. Of these, 86 (69.9%, 86/123) participants completed shop 1, 79 (64.2%, 79/123) completed shop 2, 74 (60.2%, 74/123) completed shop 3, and 60 (48.8%, 60/123) returned their till receipts (see Figure 2).

Table 2 shows the characteristics of the participants who completed at least 1 virtual shop (N=86), completed all 3 virtual shops and returned valid till receipts (N=52), and completed all 3 virtual shops and returned valid till receipts plus had a high presence score (N=25). Results show that there were no clear differences in these sets of participants, except that participants in group 3 (high presence) were somewhat younger (mean 34.0, SD 8.5 years compared to mean 38.7, SD 12.3 years for all 86 participants), and included fewer retired persons and more students.

Figure 2. Study flow diagram.

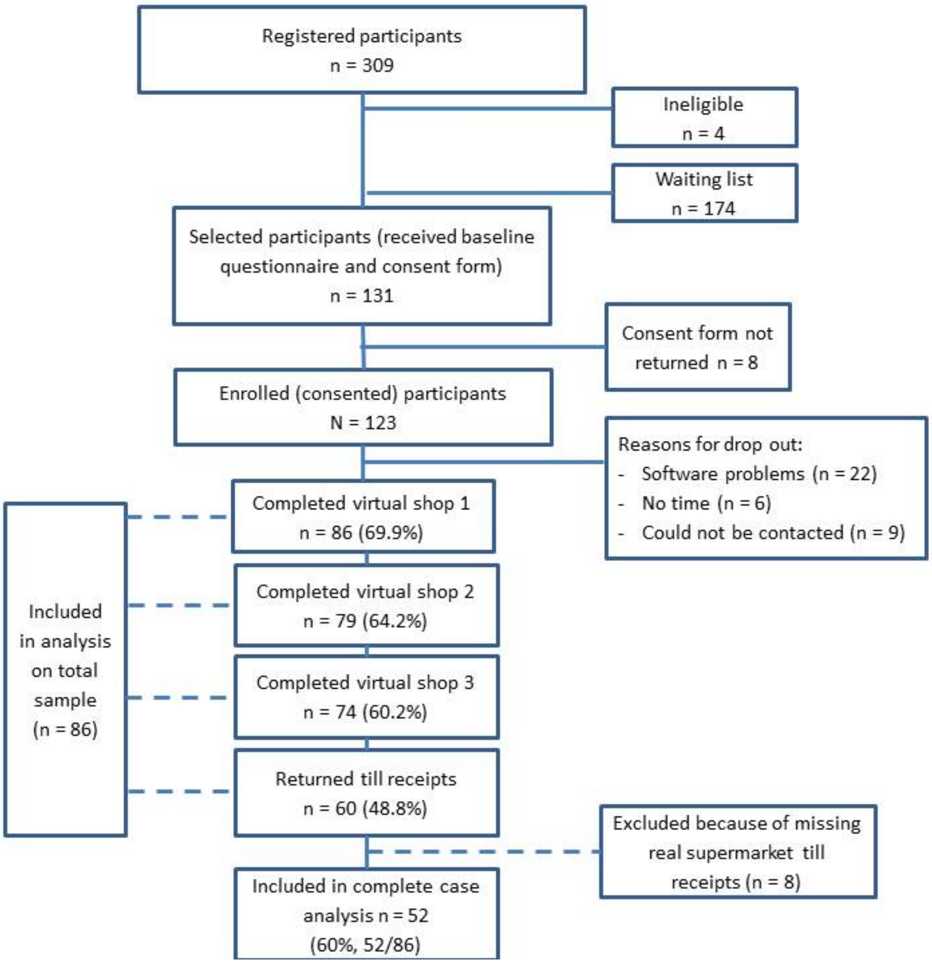


Table 2. Participant characteristics (N=86).

Participant characteristics	All participants[a] N=86	Complete cases[b] n=52	High presence[c] n=25
Age (years), mean (SD)	38.7 (12.3)	38.0 (11.2)	34.0 (8.5)
Gender (female), n (%)	66 (76.7)	41 (79)	20 (80)
Household size, n (%)			
1 person	9 (10.5)	7 (14)	3 (12)
2 persons	32 (37.2)	19 (37)	10 (40)
3 persons	15 (17.4)	9 (17)	6 (24)
≥4 persons	28 (32.6)	16 (31)	6 (16)
Missing	2 (2.3)	1 (2)	—
Ethnicity, n (%)			
NZ European	45 (52.3)	25 (48)	12 (48)
Māori	3 (3.5)	2 (4)	1 (4)
Pacific Islander	3 (3.5)	1 (2)	1 (4)
Chinese	6 (7.0)	4 (8)	1 (4)
Other (Australian, English, etc)	29 (33.7)	20 (38)	10 (36)
Responsibility for household grocery shopping, n (%)			
Mainly responsible	51 (59.3)	32 (62)	14 (56)
Mostly responsible	17 (19.8)	7 (14)	3 (12)
Shared responsibility	17 (19.8)	13 (25)	8 (32)
Missing	1 (1.2)	—	—
Highest educational qualification, n (%)			
None	4 (4.7)	2 (4)	—
Secondary school	11 (12.8)	5 (10)	3 (12)
Undergraduate	26 (30.2)	19 (37)	9 (36)
Postgraduate	34 (39.5)	20 (39)	11 (44)
Other	4 (4.7)	3 (6)	1 (4)
Declined answer	7 (8.2)	3 (6)	1 (4)
Work status, n (%)			
Full time paid work	34 (39.5)	22 (42)	12 (48)
Part time paid work	14 (16.3)	9 (17)	4 (16)
Retired	6 (7.0)	3 (6)	0 (0)
Student	11 (12.8)	6 (12)	4 (16)
Homemaker	11 (12.8)	7 (14)	4 (16)
Beneficiary	2 (2.3)	0 (0)	0 (0)
Other	7 (8.2)	5 (10)	1 (4)
Missing	1 (1.2)	—	—
Total household income before tax [d] (NZ $), n (%)			
<$20,000	5 (5.8)	4 (8)	2 (8)
$20,001-$40,000	11 (12.8)	7 (13)	3 (12)
$40,001-$60,000	11 (12.8)	5 (10)	1 (4)
$60,001-$80,000	16 (18.6)	9 (17)	7 (28)
≥$80,001	39 (45.4)	20 (39)	11 (44)

Participant characteristics	All participants[a] N=86	Complete cases[b] n=52	High presence[c] n=25
Don’t know/declined answer	4 (4.7)	2 (4)	1 (4)
Food purchases at supermarket, n (%)			
All food	9 (10.5)	6 (12)	4 (16)
Most food	62 (72.1)	36 (70)	17 (68)
Half of food	12 (14.0)	9 (17)	3 (12)
Some food	2 (2.4)	1 (2)	1 (4)
Missing	1 (1.2)	—	—
Grocery shopping frequency, n (%)			
< Once a week	10 (11.6)	5 (10)	2 (8)
Once a week	31 (36.0)	16 (31)	8 (32)
Twice a week	29 (33.7)	22 (42)	10 (40)
≥3 times a week	15 (17.5)	9 (17)	5 (20)
Missing	1 (1.2)	—	—
Frequency of using a computer, n (%)			
Daily	83 (96.5)	51 (98)	24 (96)
Several days per week	2 (2.3)	1 (2)	1 (4)
Missing	1 (1.2)	—	—
Internet at home (yes), n (%)	85 (98.8)	51 (98)	24 (96)
Type of operating system used for this study, n (%)			
Windows XP	11 (12.8)	6 (12)	4 (16)
Windows 7	41 (47.7)	26 (50)	11 (44)
Windows (don’t know version)	20 (23.3)	11 (21)	7 (28)
MacOS X 10.5/10.6	5 (5.8)	3 (6)	1 (4)
MacOS X 10.7/10.8	6 (7.0)	5 (10)	1 (4)
MacOS X (don’t know version)	2 (2.3)	1 (2)	1 (4)
Missing	1 (1.2)	—	—

[a] Includes all participants who completed at least 1 shop in the Virtual Supermarket.

[b] Includes all participants who completed all 3 shops in the Virtual Supermarket and returned valid till receipts.

[c] Includes participants who completed all 3 shops in the Virtual Supermarket and returned valid till receipts and had a high presence score.

[d] Median annual income in New Zealand from wages and salaries was NZ $43,888 in June 2013 (per person)

Descriptive Analysis

Figures 3 and 4 show the proportion of money spent / proportion of items purchased for each of the 18 food groups, respectively, for both the virtual and the real supermarket and for each of the 3 shopping occasions (n=86). Results suggest that the purchasing patterns across the food groups were comparable. The 4 food groups with the highest relative expenditures were the same for the virtual and real supermarkets (all 3 shops): fruit and vegetables (14.3% Virtual Supermarket and 17.4% real supermarket), bread and bakery (10.0% Virtual Supermarket and 8.2% real supermarket), dairy (19.1% Virtual Supermarket and 12.6% real supermarket), and meat and fish (16.5% Virtual Supermarket and 16.8% real supermarket) (Table 3). Overall, total mean expenditures in the Virtual Supermarket were NZ $84.96 (SD 46.88 compared to NZ $125.15 (SD 74.15) in the real supermarket. On average, participants set their budget in the Virtual Supermarket at NZ $121.19 (SD 65.01) and spent 71.4% (SD 25.6) of this budget.

Table 3. Means and differences for Virtual Supermarket and real supermarket purchases for the 18 food categories (complete and valid cases, n=52).

Food category	% Total expenditures				% Total number of items			
	Means		Difference estimate	*P*	Means		Difference estimate	*P*
	Virtual estimate[a]	Real estimate			Virtual estimate	Real estimate		
Alcoholic beverages	5.15	3.47	1.68	.13	1.96	1.46	0.50	.43
Baby care	0.83	0.53	0.30	.43	1.21	0.96	0.26	.54
Bread and bakery	10.0	8.15	1.89	.02	11.7	10.3	1.44	.08
Baking and cooking	1.10	0.56	0.54	.16	0.98	0.85	0.39	.17
Beverages	5.03	5.92	–0.89	.30	4.63	5.32	–0.69	.21
Cereal and cereal products	5.64	7.14	–1.50	.06	5.46	7.58	–2.13	.01
Convenience foods	1.70	3.07	–1.38	.10	1.68	3.35	–1.70	.02
Dairy	19.1	12.6	6.49	<.001	13.6	11.4	2.21	.04
Deli and chilled foods	0.31	0.07	0.25	.15	0.28	0.07	0.21	.29
Edible oils and emulsions	3.28	3.16	0.12	.84	2.46	2.35	0.11	.81
Eggs	3.78	2.15	1.63	.01	2.94	1.50	1.44	.003
Fish and fish products	2.82	2.44	0.38	.42	2.88	2.65	0.23	.69
Fruit and vegetables fresh	14.3	17.4	–3.12	.04	28.9	22.9	5.94	.002
Fruit and vegetables other	5.46	7.36	–1.90	.009	6.57	8.64	–2.07	.02
Fresh meat, meat products, fresh fish	16.5	16.8	–0.31	.85	9.64	9.77	–0.13	.89
Other miscellaneous	0.05	0.42	–0.36	.07	0.07	0.35	–0.28	.12
Sauces, spreads, and seasonings	3.38	4.56	–1.18	.05	3.29	4.79	–1.50	.01
Snack foods	1.58	4.17	–2.59	<.001	1.83	5.88	–4.05	<.001

[a] Results from repeated measures mixed models fitted to evaluate the percentage of price (outcome 1) and percentage of items (outcome 2), respectively, with supermarket type (virtual or real) and shop (1, 2, or 3) as fixed effects, registration number as random effect, and shop as repeated effect.

Figure 3. Proportion (%) expenditures for the 18 foods groups for the virtual and real shopping occasions (raw data) (n=86).

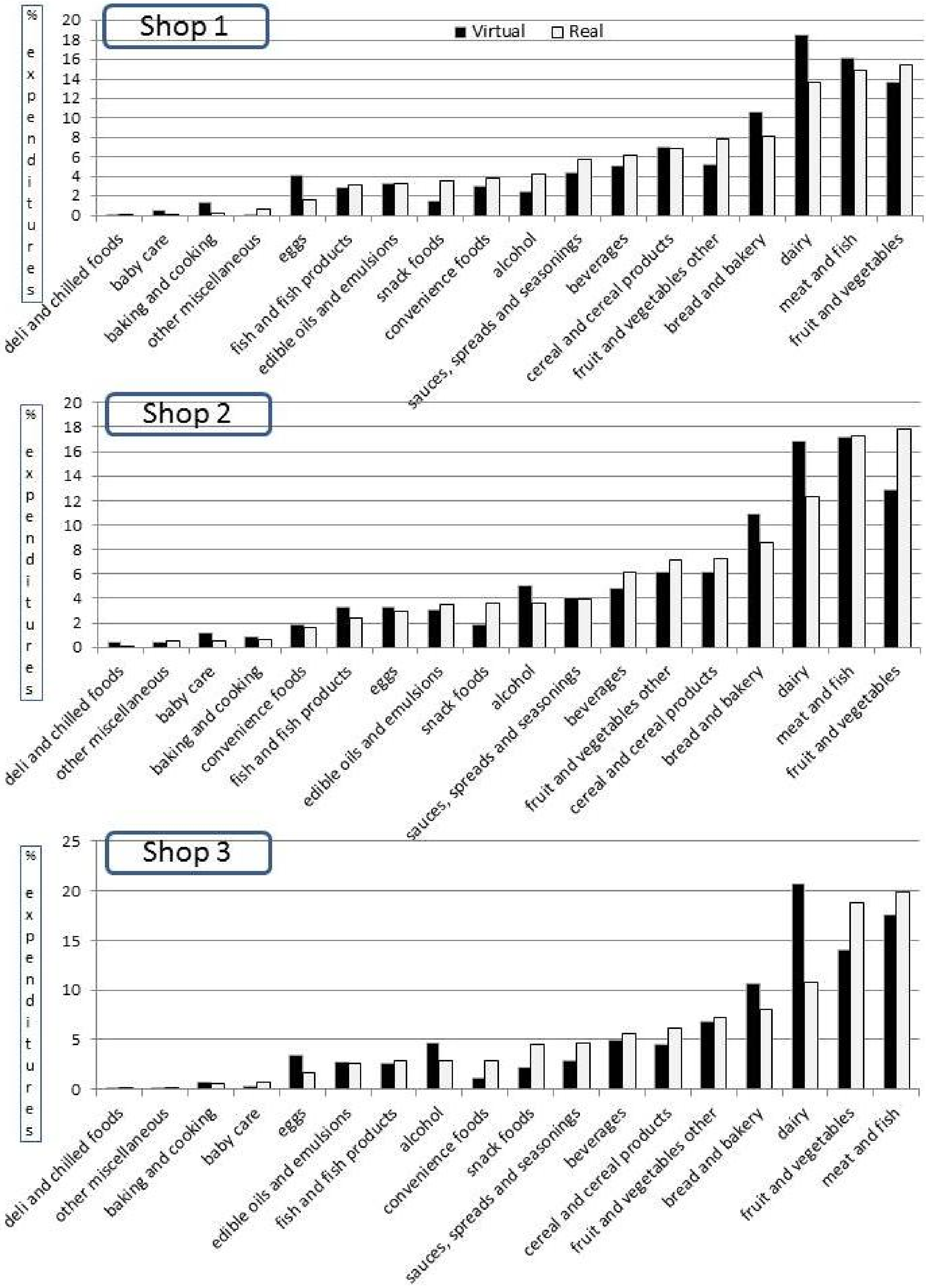
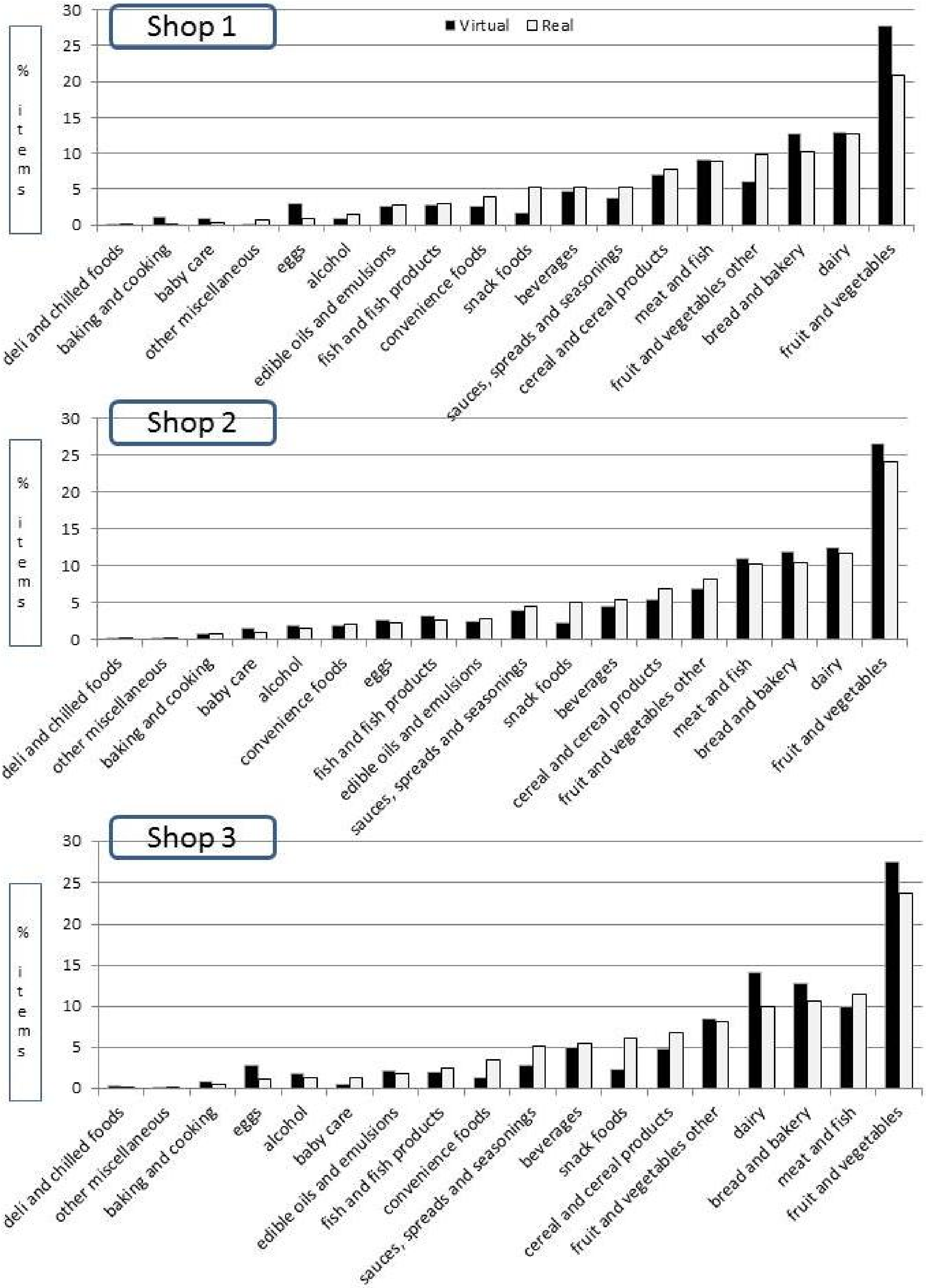


Figure 4. Proportion (%) of items purchased for the 18 foods groups for the virtual and real shopping occasions (raw data) (n=86).



Difference Between Virtual and Real Purchases

Table 3 shows the estimated means and difference for the virtual and real purchases for all 18 food categories and separately as a proportion of total expenditures (outcome 1) and proportion of total items (outcome 2) for the complete cases (n=52). Results for the total sample (N=86) are shown in Multimedia Appendix 1 and show comparable outcomes to the results presented subsequently. Mixed models revealed that the effect of shopping occasion (shop 1, 2, or 3) was only significant for the convenience foods category. This means that the similarity between virtual and real purchases was stable over time and, therefore, the results are presented at an aggregate level. Six food categories showed significant differences in purchases between the virtual and real shop, including bread and bakery (difference estimate 1.9%, *P*=.02); dairy (6.5%, *P*<.001); fruit and vegetables, fresh (3.1%, *P*=.04); fruit and vegetables, other (1.9%, *P*=.009); snack foods (2.6%, *P*<.001); and eggs (1.6%, *P*=.01). Similar results were observed for difference in proportion of items purchased.

Focus and Presence

Previous research indicated that people with a higher focus might perceive stronger feelings of presence in virtual environments. Outcomes of the linear mixed models confirmed that focus was strongly positively associated with perceived presence in the Virtual Supermarket (estimate 1.05, *P*<.001).

With regard to presence, results showed that a large majority of participants (≥95% across the 3 shops) reported a medium-to-high feeling of presence in the Virtual Supermarket (Table 4). Participants scored particularly high for the subscales adaptation/immersion (ie, perceiving oneself to be enveloped by, included in, and interacting with an environment) and interface quality. According to theory, we would expect that people with a higher presence would perform better in the Virtual Supermarket compared to people with a lower perception of presence. Sensitivity analysis on a subset of participants with a high presence score (n=25) revealed greater similarity in proportional expenditures between virtual and real purchasing behavior. The total number of categories with significant differences between virtual and real purchases reduced from 6 to 4 in this subset of participants (the categories “fruit and vegetables” and “eggs” showed no longer significant differences).

Table 4. Mean total presence score[a] and subscales for all participants who completed at least 1 virtual shopping occasion (n=83[b]).

Presence score and subscales for 3 groups	Mean (SD)	n (%)
Total presence (range 29-203)	104 (22.0)	
Low (0-67)		3 (4)
Medium (68-134)		73 (88)
High (≥134)		7 (8)
Subscale 1: involvement (range 12-84)	40.0 (11.1)	
Low (0-28)		12 (15)
Medium (29-56)		67 (81)
High (≥57)		4 (5)
Subscale 2: sensory fidelity (range 6-42)	18.0 (4.8)	
Low (0-14)		19 (23)
Medium (15-28)		62 (75)
High (≥29)		2 (2)
Subscale 3: adaptation/immersion (range 8-56)	31.2 (6.9)	
Low (0-18)		1 (1)
Medium (19-37)		64 (77)
High (≥38)		18 (22)
Subscale 4: interface quality (range 3-21)	14.9 (3.2)	
Low (0-7)		1 (1)
Medium (8-14)		33 (40)
High (≥15)		49 (59)
Focus (range 7-49)	29.7 (5.7)	
Low (0-16)		—
Medium (17-32)		51 (61)
High (≥33)		32 (39)

[a] Mean represents the typical value for each participant over all recorded virtual shops.

[b] Questionnaire data for 3 participants were missing due to technical issues.

Discussion

This study shows that consumer purchasing patterns in the Virtual Supermarket were comparable to those in real supermarkets and confirms that the Virtual Supermarket is a valid tool to measure food purchasing behavior in a supermarket setting. Furthermore, a large majority of participants experienced a medium-to-high sense of presence in the Virtual Supermarket. The categories "fruits and vegetables" and "dairy" showed the strongest difference between virtual and real purchases. Results will be useful to guide further improvement of the software.

This validation study examined differences in purchasing patterns between virtual and real shopping occasions across a total of 18 different food categories. Results demonstrate that the 4 food groups with the highest relative expenditures were the same for the virtual and real shopping occasions (eg, fresh fruit and vegetables, bread and bakery, dairy, and meat and fish). Likewise, the 4 food groups with the lowest proportional expenditures were the same (eg, other miscellaneous, deli and chilled foods, baking and cooking goods, and baby care products). These results indicate that the overall purchasing patterns were similar in the virtual and real supermarket and that proportional purchases within food groups were comparable.

One key limitation of the Virtual Supermarket is that participants did not use real money and, therefore, could spend more than they would in real life. For this reason, participants were asked to set a realistic shopping budget and the software did not allow them to overspend. Results showed no patterns of overspending in the Virtual Supermarket compared to the real supermarket; for some food categories, people did spend more in the Virtual Supermarket (dairy, bread and bakery, and eggs) but for others they spent less (fruit and vegetables, and snack foods). Unexpectedly, total absolute expenditures were substantially lower in the Virtual Supermarket compared to the real supermarket. On average, participants spent approximately 70% of their shopping budget in the Virtual Supermarket (which is reasonable); however, in approximately 10% of the shopping occasions, participants spent less than 30% of their budget. One plausible explanation for the relatively low expenditures is that the Virtual Supermarket holds no sales promotions. A 2002 study by Aguirregabiria [28] revealed that typical sales promotions (which are approximately a 20% discount) increase weekly sales rates by 500%. In addition, choices are often influenced by other in-store promotions, such as end-of-aisle and merchandising displays [29], none of which are available in our Virtual Supermarket.

We could include promotions in future versions of the Virtual Supermarket; however, when designing RCTs it is important to realize that these features might influence the outcome of interest. One key advantage of the Virtual Supermarket compared to real supermarkets is that it enables creation of a "neutral" shopping environment with completely controllable experimental and control conditions enabling effective testing of the effects of changing only one element (eg, price) without interference of other effects (eg, sales signs). Also, it can be used nationwide because it does not require participants to come to a specific research location. Depending on the specific study aims, it is important to consider the balance between making the Virtual Supermarket as realistic as possible (external validity) and designing it to deliver a valid study outcome (internal validity). A major advantage of the Virtual Supermarket is that it is modifiable and researchers can design the Virtual Supermarket in accordance with specific study needs.

In addition to overall shopping patterns, this study compared proportional purchases within 18 food categories. Results demonstrated that 6 of the food categories differed significantly in proportion of total expenditures between the virtual and real shops. These categories were bread and bakery; dairy; fruit and vegetables, fresh; fruit and vegetables, other; snack foods; and eggs. Although this finding is important, it needs to be interpreted with some caution. Most importantly, we made multiple comparisons between groups, which increased the likeliness of finding significant results. A more conservative level of significance of P=.01 might be appropriate, in which case only 3 categories (dairy, processed fruit and vegetables, and snack foods) would have shown relevant significant differences. Furthermore, our focus was on proportional purchases, meaning that if one food category showed clear differences, other food categories would be affected as well (eg, if participants spent 17% on meat in the Virtual Supermarket there was only 83% left for other categories). The reason why we focused on relative purchases and not absolute purchases is that the prices in the Virtual Supermarket and the real supermarket were expected to be different due to promotional and seasonality effects in the real supermarket. Finally, it is important to consider that our findings are quite conservative. Our conclusions are based on comparing shopping behavior in 2 different environments. However, it also can be expected that 2 shops in the same environment can be slightly different (eg, people do not buy exactly the same each time). Our analysis did not account for this "natural" variation in shopping behavior, which further strengthens our finding that the differences between the Virtual Supermarket and the real supermarket were relatively minor.

Keeping this in mind, food categories that showed the most important differences between the virtual and real shopping occasions were fresh fruit and vegetables, dairy, and snack foods. For fresh fruit and vegetables, participants spent significantly more in real life compared to the virtual shops. One important explanation for this finding is that many participants misunderstood the interface in the Virtual Supermarket for purchasing loose items. After clicking on a fruit or vegetable product, participants were asked to enter the number of grams they wanted to buy (eg, 100 grams of apples). However, we observed that a large number of participants (n=29) misinterpreted this question and entered the number of items they wanted to buy (eg, 2 apples). The software registered this automatically as grams, meaning that the till receipt would show 2 grams of apples. This error caused a significant underrepresentation of fruit and vegetable expenditures in the Virtual Supermarket and if these participants were excluded from analysis the proportional expenditures on fruit and vegetables increased from mean of approximately 14% (SD 14) to a mean of approximately 15% (SD 15). This effect might also partly explain the observed underspending in the Virtual

Supermarket as mentioned previously. Therefore, improving the interface for loose items will be a key priority when updating the Virtual Supermarket. Looking at snack foods, the relative underspending within this category in the Virtual Supermarket could be due to a number of reasons. First, snack foods are normally heavily promoted which leads to increased sales [29-31]. This effect was not captured in the Virtual Supermarket. Also, supermarkets tend to provide more shelf space for snacks foods, which is expected to increase sales [32]. The Virtual Supermarket used a representative product selection based on the number of different types of products/brands available for sale in each food category in NZ (eg, different types of soft drinks), not accounting for the exact number of individual products that is normally displayed on the shelves (eg, supermarkets often display a large number of the same Coca-Cola bottles covering a whole aisle; the Virtual Supermarket did not account for this). To improve the resemblance with real supermarkets, it might be important to measure shelf space for all products in real supermarkets and design the Virtual Supermarket accordingly. However, there is a large variation in design between different supermarkets making it hard to determine the best representation; also, this requires cooperation from supermarkets which might be difficult to achieve.

Another limitation of this study was the relatively large number of participants that dropped out due to software problems. In particular, a large number of participants reported that the software could not be run on their (mostly older) computers. Since completion of this validation study, we have found a solution to this by improving the way the software processes the individual product images. Therefore, for future studies, the NZ Virtual Supermarket will be more compatible with older computers and laptops. Nevertheless, the high dropout could bias our results and participants that completed the study might have been better using the software than an average user. Another problem was that we were not able to recruit an ethnically diverse sample and, therefore, are not able to draw conclusions on the acceptability and usability of the Virtual Supermarket among Māori (indigenous) and Pacific people. To ensure diversity in future studies, it would be important to use targeted recruitment strategies for these populations. Nevertheless, our sample was relatively diverse with regard to other characteristics, including education level, age, work status, and income, and there were no signs of differential dropout.

To our knowledge, this is one of the first studies making a direct comparison between virtual and real shopping behavior. One other example is a 2004 study by Grewe et al [33] who validated a virtual reality (VR) shopping task to assess real-life memory and spatial navigation in patients with epilepsy. The authors found that the VR tool was valid and generalizable compared to a real-life shopping task. Furthermore, Sharpe et al [34] validated fast food and soft drink purchases made in an online simulated road trip against actual purchasing and consumption of a lunch meal from McDonalds 1 week later. This study found strong evidence that virtual and real purchases were similar and that participants were consistent in the type of foods they chose in the virtual task [34]. In other health-related domains, there is more evidence for the realism and performance of virtual environments. For example, VR is frequently used to improve surgical technical skills [35], as a therapeutic instrument in clinical psychology [36,37], and in the treatment of several phobias [38]. Also, in the field of computer sciences, there is a wealth of research into human computer interaction [39]. Linking this evidence to the Virtual Supermarket, there is good reason to assume that people experience a high presence because of the high realism (eg, it requires very little imagination to recognize the software as a supermarket), the realistic movement (pushing a shopping trolley, 3D features), and the fact that it is easy to locate oneself due to clear landmarks (eg, aisles) [39]. One way to further improve presence is through the use of avatars. People tend to feel a higher feeling of presence when there is a virtual representation of oneself in the virtual world and when other users recognize them (other avatars) [39]. It would be interesting to explore the use of avatars further and to test the impact of social interactions in the Virtual Supermarket.

Another way to improve the feeling of presence is by improving software graphics and functionality. With improving technology, virtual environments are becoming increasingly realistic and we aim to improve the Virtual Supermarket alongside improving technology. Although we realize that virtual worlds will never be the same as real life, they do provide a clear opportunity to conduct research that would otherwise be nearly impossible. In addition, the Virtual Supermarket offers the opportunity to provide much needed independent rigorous experimental evidence to inform (controversial) food policy [40].

This study shows that the Virtual Supermarket is a valid and useful tool to study food-purchasing behavior. The results of this validation will provide valuable guidance regarding further improvements to the software and to the optimization of the internal and external validity of this innovative methodology.

Acknowledgments

This work was supported by a University of Auckland Faculty Research Development Fund grant. The lead author is supported by a Fellowship from the National Heart Foundation (grant 1570). We acknowledge Professor Tony Blakely for his help in critically reviewing the manuscript and our research assistant Rita George for her help with the completion of study procedures. The Virtual Supermarket was developed by SURFsara Amsterdam and funded by the VU University Amsterdam.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Means and difference for Virtual Supermarket and real supermarket purchases for the 18 food categories (total sample, n=86).

[PDF File (Adobe PDF File), 72KB - jmir_v17i4e107_app1.pdf]

References

1. Gortmaker SL, Swinburn BA, Levy D, Carter R, Mabry PL, Finegood DT, et al. Changing the future of obesity: science, policy, and action. Lancet 2011 Aug 27;378(9793):838-847 [FREE Full text] [doi: 10.1016/S0140-6736(11)60815-5] [Medline: 21872752]
2. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. Lancet 2011 Aug 27;378(9793):804-814. [doi: 10.1016/S0140-6736(11)60813-1] [Medline: 21872749]
3. Roberto CA, Pomeranz JL, Fisher JO. The need for public policies to promote healthier food consumption: A comment on Wansink and Chandon (2014). Journal of Consumer Psychology 2014 Jul;24(3):438-445. [doi: 10.1016/j.jcps.2014.03.001]
4. Costa JC, Claro RM, Martins APB, Levy RB. Food purchasing sites. Repercussions for healthy eating. Appetite 2013 Nov;70:99-103. [doi: 10.1016/j.appet.2013.06.094] [Medline: 23835229]
5. Ni Mhurchu C, Vandevijvere S, Waterlander W, Thornton LE, Kelly B, Cameron AJ, INFORMAS. Monitoring the availability of healthy and unhealthy foods and non-alcoholic beverages in community and consumer retail food environments globally. Obes Rev 2013 Oct;14 Suppl 1:108-119. [doi: 10.1111/obr.12080] [Medline: 24074215]
6. Hawkes C. Dietary implications of supermarket development: a global perspective. Development Policy Review 2008;26(6):657-692. [doi: 10.1111/j.1467-7679.2008.00428.x]
7. Vorley B, UK Food Group. Food, Inc: Corporate Concentration from Farm to Consumer. London: International Institute for Environment and Development; 2003. URL: http://www.ukfg.org.uk/docs/UKFG-Foodinc-Nov03.pdf [accessed 2014-08-10] [WebCite Cache ID 6RjXJbnBB]
8. Health Sponsorship Council. In fact: Shopping behaviours of New Zealand households. 2012 Feb. URL: http://archive.hsc.org.nz/sites/default/files/publications/In%20Fact-Shopping%20behaviours-fnl-120301.pdf [accessed 2014-08-10] [WebCite Cache ID 6RjX9tpQz]
9. Glanz K, Hoelscher D. Increasing fruit and vegetable intake by changing environments, policy and pricing: restaurant-based research, strategies, and recommendations. Prev Med 2004 Sep;39 Suppl 2:S88-S93. [doi: 10.1016/j.ypmed.2004.03.002] [Medline: 15313077]
10. Escaron AL, Meinen AM, Nitzke SA, Martinez-Donate AP. Supermarket and grocery store-based interventions to promote healthful food choices and eating practices: a systematic review. Prev Chronic Dis 2013;10:E50 [FREE Full text] [doi: 10.5888/pcd10.120156] [Medline: 23578398]
11. Ni Mhurchu C, Blakely T, Jiang Y, Eyles HC, Rodgers A. Effects of price discounts and tailored nutrition education on supermarket purchases: a randomized controlled trial. Am J Clin Nutr 2010 Mar;91(3):736-747 [FREE Full text] [doi: 10.3945/ajcn.2009.28742] [Medline: 20042528]
12. Waterlander WE, de Boer MR, Schuit AJ, Seidell JC, Steenhuis IHM. Price discounts significantly enhance fruit and vegetable purchases when combined with nutrition education: a randomized controlled supermarket trial. Am J Clin Nutr 2013 Apr;97(4):886-895 [FREE Full text] [doi: 10.3945/ajcn.112.041632] [Medline: 23446898]
13. Hersey JC, Wohlgenant KC, Arsenault JE, Kosa KM, Muth MK. Effects of front-of-package and shelf nutrition labeling systems on consumers. Nutr Rev 2013 Jan;71(1):1-14. [doi: 10.1111/nure.12000] [Medline: 23282247]
14. Stanford DD. Bloomberg Business Week. 2012 Mar 14. Anti-obesity soda tax fails as lobbyists spend millions URL: http://www.bloomberg.com/news/articles/2012-03-13/anti-obesity-soda-tax-fails-as-lobbyists-spend-millions-retail [WebCite Cache ID 6Xp5gODA5]
15. Nichols S, Haldane C, Wilson JR. Measurement of presence and its consequences in virtual environments. International Journal of Human-Computer Studies 2000 Mar;52(3):471-491. [doi: 10.1006/ijhc.1999.0343]
16. Waterlander WE, Scarpa M, Lentz D, Steenhuis IHM. The virtual supermarket: an innovative research tool to study consumer food purchasing behaviour. BMC Public Health 2011;11:589 [FREE Full text] [doi: 10.1186/1471-2458-11-589] [Medline: 21787391]
17. Nederkoorn C, Guerrieri R, Havermans RC, Roefs A, Jansen A. The interactive effect of hunger and impulsivity on food intake and purchase in a virtual supermarket. Int J Obes (Lond) 2009 Aug;33(8):905-912. [doi: 10.1038/ijo.2009.98] [Medline: 19546869]
18. Josman N, Kizony R, Hof E, Goldenberg K, Weiss PL, Klinger E. Using the virtual action planning-supermarket for evaluating executive functions in people with stroke. J Stroke Cerebrovasc Dis 2014;23(5):879-887. [doi: 10.1016/j.jstrokecerebrovasdis.2013.07.013] [Medline: 24008131]
19. Van Kleef E. Food Intake Control. 23 Mar 2012. Have a look at the new virtual supermarket of Wageningen University URL: http://foodintakecontrol.blogspot.co.nz/2012/03/have-look-at-new-virtual-supermarket-of.html [accessed 2014-08-11] [WebCite Cache ID 6RjcrYdcK]

20. Witmer BG, Jerome CJ, Singer MJ. The factor structure of the presence questionnaire. Presence: Teleoperators and Virtual Environments 2005 Jun;14(3):298-312. [doi: 10.1162/105474605323384654]
21. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. Presence 1998 Jun;7(3):225-240. [doi: 10.1162/105474698565686]
22. Waterlander WE, Ni Mhurchu C, Steenhuis IHM. Effects of a price increase on purchases of sugar sweetened beverages. Results from a randomized controlled trial. Appetite 2014 Jul;78:32-39. [doi: 10.1016/j.appet.2014.03.012] [Medline: 24667153]
23. Waterlander WE, Steenhuis IHM, de Boer MR, Schuit AJ, Seidell JC. Introducing taxes, subsidies or both: the effects of various food pricing strategies in a web-based supermarket randomized trial. Prev Med 2012 May;54(5):323-330 [FREE Full text] [doi: 10.1016/j.ypmed.2012.02.009] [Medline: 22387008]
24. Waterlander WE, Ni Mhurchu C, Steenhuis IH. The use of virtual reality in studying complex interventions in our every-day food environment. In: Tang X, editor. Virtual Reality Human Computer Interaction. Rijeka, Croatia: InTech; 2012:229-260.
25. Bioinformatics. The Virtual Supermarket: an innovative research tool to study food purchasing behaviour URL: http://www.bioinformatics.org/supermarket/index.html [accessed 2015-04-16] [WebCite Cache ID 6Xp68CHc5]
26. Woodward E, Eyles H, Ni Mhurchu C. Key opportunities for sodium reduction in New Zealand processed foods. Aust N Z J Public Health 2012 Feb;36(1):84-89. [doi: 10.1111/j.1753-6405.2012.00829.x] [Medline: 22313711]
27. Retail World Annual Report. Rydalmere, NSW: Retail Media; 2010.
28. Aguirregabiria V. Sales promotions in supermarkets: Estimating their effects on profits and consumer welfare. Working paper. Boston: Boston University; 2002. URL: http://economics.yale.edu/sites/default/files/files/Workshops-Seminars/Industrial-Organization/aguirre-030327.pdf [accessed 2014-08-10] [WebCite Cache ID 6RjXcjSmW]
29. Glanz K, Bader MDM, Iyer S. Retail grocery store marketing strategies and obesity: an integrative review. Am J Prev Med 2012 May;42(5):503-512. [doi: 10.1016/j.amepre.2012.01.013] [Medline: 22516491]
30. Nestle M. Conclusion: The politics of food choice. In: Goldstein E, editor. Food Politics: How the Food Industry Influences Nutrition and Health. Berkeley: University of California Press; 2007:358-374.
31. Thornton LE, Cameron AJ, McNaughton SA, Worsley A, Crawford DA. The availability of snack food displays that may trigger impulse purchases in Melbourne supermarkets. BMC Public Health 2012;12:194 [FREE Full text] [doi: 10.1186/1471-2458-12-194] [Medline: 22420759]
32. Thornton LE, Cameron AJ, McNaughton SA, Waterlander WE, Sodergren M, Svastisalee C, et al. Does the availability of snack foods in supermarkets vary internationally? Int J Behav Nutr Phys Act 2013;10:56 [FREE Full text] [doi: 10.1186/1479-5868-10-56] [Medline: 23672409]
33. Grewe P, Lahr D, Kohsik A, Dyck E, Markowitsch HJ, Bien CG, et al. Real-life memory and spatial navigation in patients with focal epilepsy: ecological validity of a virtual reality supermarket task. Epilepsy Behav 2014 Feb;31:57-66. [doi: 10.1016/j.yebeh.2013.11.014] [Medline: 24361763]
34. Sharpe K, Staelin R, Huber J. Using extremeness aversion to fight obesity: policy implications of context dependent demand. J Consum Res 2008 Oct;35(3):406-422. [doi: 10.1086/587631]
35. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002 Oct;236(4):458-63; discussion 463. [doi: 10.1097/01.SLA.0000028969.51489.B4] [Medline: 12368674]
36. Riva G. Virtual reality in psychotherapy: review. Cyberpsychol Behav 2005 Jun;8(3):220-30; discussion 231. [doi: 10.1089/cpb.2005.8.220] [Medline: 15971972]
37. Cesa GL, Manzoni GM, Bacchetta M, Castelnuovo G, Conti S, Gaggioli A, et al. Virtual reality for enhancing the cognitive behavioral treatment of obesity with binge eating disorder: randomized controlled study with one-year follow-up. J Med Internet Res 2013;15(6):e113 [FREE Full text] [doi: 10.2196/jmir.2441] [Medline: 23759286]
38. Wiederhold BK, Gevirtz R, Wiederhold MD. Fear of flying: a case report using virtual reality therapy with physiological monitoring. Cyberpsychology & Behavior 1998 Jan;1(2):97-103. [doi: 10.1089/cpb.1998.1.97]
39. Nash EB, Edwards GW, Thompson JA, Barfield W. A review of presence and performance in virtual environments. International Journal of Human-Computer Interaction 2000 May;12(1):1-41. [doi: 10.1207/S15327590IJHC1201_1]
40. Ludwig DS, Nestle M. Can the food industry play a constructive role in the obesity epidemic? JAMA 2008 Oct 15;300(15):1808-1811. [doi: 10.1001/jama.300.15.1808] [Medline: 18854543]

Abbreviations

RCT: randomized controlled trials
SMS: short message service
VR: virtual reality

Edited by G Eysenbach; submitted 11.08.14; peer-reviewed by P Bordnick, S Boucher; comments to author 23.10.14; revised version received 25.11.14; accepted 13.12.14; published 28.04.15.

Please cite as:
Waterlander WE, Jiang Y, Steenhuis IHM, Ni Mhurchu C
Using a 3D Virtual Supermarket to Measure Food Purchase Behavior: A Validation Study
J Med Internet Res 2015;17(4):e107
URL: http://www.jmir.org/2015/4/e107/
doi:10.2196/jmir.3774
PMID:25921185

Original Paper

Identifying Key Hospital Service Quality Factors in Online Health Communities

Yuchul Jung[1], PhD; Cinyoung Hur[2], MS Comp Sc; Dain Jung[3], BSC; Minki Kim[4], PhD

[1]Korea Institute of Science and Technology Information, Daejeon, Republic Of Korea

[2]Electronics and Telecommunications Research Institute, Daejeon, Republic Of Korea

[3]Korea Advanced Institute of Science and Technology, Department of Business and Technology Management, Daejeon, Republic Of Korea

[4]Korea Advanced Institute of Science and Technology, College of Business, Seoul, Republic Of Korea

Corresponding Author:
Minki Kim, PhD
Korea Advanced Institute of Science and Technology
College of Business
S304, KAIST Business School
85 Hoegiro, Dongdaemun-gu
Seoul, 130-722
Republic Of Korea
Phone: 82 29583512
Fax: 82 29583620
Email: minki.kim@kaist.ac.kr

Abstract

Background: The volume of health-related user-created content, especially hospital-related questions and answers in online health communities, has rapidly increased. Patients and caregivers participate in online community activities to share their experiences, exchange information, and ask about recommended or discredited hospitals. However, there is little research on how to identify hospital service quality automatically from the online communities. In the past, in-depth analysis of hospitals has used random sampling surveys. However, such surveys are becoming impractical owing to the rapidly increasing volume of online data and the diverse analysis requirements of related stakeholders.

Objective: As a solution for utilizing large-scale health-related information, we propose a novel approach to identify hospital service quality factors and overtime trends automatically from online health communities, especially hospital-related questions and answers.

Methods: We defined social media–based key quality factors for hospitals. In addition, we developed text mining techniques to detect such factors that frequently occur in online health communities. After detecting these factors that represent qualitative aspects of hospitals, we applied a sentiment analysis to recognize the types of recommendations in messages posted within online health communities. Korea’s two biggest online portals were used to test the effectiveness of detection of social media–based key quality factors for hospitals.

Results: To evaluate the proposed text mining techniques, we performed manual evaluations on the extraction and classification results, such as hospital name, service quality factors, and recommendation types using a random sample of messages (ie, 5.44% (9450/173,748) of the total messages). Service quality factor detection and hospital name extraction achieved average F1 scores of 91% and 78%, respectively. In terms of recommendation classification, performance (ie, precision) is 78% on average. Extraction and classification performance still has room for improvement, but the extraction results are applicable to more detailed analysis. Further analysis of the extracted information reveals that there are differences in the details of social media–based key quality factors for hospitals according to the regions in Korea, and the patterns of change seem to accurately reflect social events (eg, influenza epidemics).

Conclusions: These findings could be used to provide timely information to caregivers, hospital officials, and medical officials for health care policies.

(J Med Internet Res 2015;17(4):e90) doi:10.2196/jmir.3646

KEYWORDS

hospital service factors; online health communities; social media-based key quality factors for hospitals; recommendation type classification; quality factor analysis; healthcare policy

Introduction

Patients and their caregivers increasingly use the Internet for health care information sharing and support [1]. In particular, they tend to obtain health information and share health experiences through social media services, such as Wikipedia, Facebook, online forums, and message boards [2]. Overall, social media apps have the potential to enhance health equity: interaction through social media can increase social support and feelings of connectedness and lead to a sense of empowerment in patients. Such interaction also increases information sharing, which in turn leads to a more patient-centric experience. For this reason, researchers studying health communication among patients have also focused on online user-generated data on health, for instance physician rating websites [3], blogs with health-related keywords [4], free-text comments about hospitals on websites [5], and message boards in online health communities [6].

Hospital quality is among the most interesting health information to patients, and many studies conducting hospital quality analysis have relied heavily on random sampling surveys [7,8]. However, the practicality of random sampling approaches is limited because they require significant human effort and are thus labor intensive and time consuming. An Internet-based survey is one alternative, but specially designed questionnaires are often limited to paid users. In addition, it becomes more difficult to apply traditional statistical approaches to health-related information owing to the tremendous volume of health-related information generated daily. Surveys regarding hospitals reflect users' opinions on the hospital's service quality and patient satisfaction, but social media can deliver the same results more readily.

Social media is a valuable resource when deciding which hospital to choose. For instance, messaging systems like Twitter and Facebook allow words and phrases to be searched using algorithms to find positive or negative sentiments [9]. Recently, a study [10] showed that a hospital's Facebook page could be a quality indicator. They found that the number of “likes” on a hospital page was an important indicator not only of patient satisfaction, but also of quality of patient care. In particular, based on observation of 40 hospitals in New York, a correlation was found between Facebook “likes” and patient recommendations as well as declines in 30-day mortality rates (ie, for every percentage point drop in the 30-day morality rate, there were nearly 100 more Facebook “likes”). Although it can be difficult to generalize much from correlations, the more important point is that widely accessible social media can be a meaningful indicator of hospital quality.

To obtain useful information from online communities, several automatic approaches have recently been reported. Investigations on various social media such as blogs, bookmarks, communities, and forums have applied topic modeling [11]. Different topic modeling strategies were considered depending on the type of social media because each social media has different document types and metadata. For example, to find an online community expert, not only authorship but also the volume of social feedback, such as comments, likes, and shares are key factors for estimating expertise. Recent studies [12-14] have used quantitative summaries of user-generated content such as overall valence and volume of user review ratings to represent user opinions. In contrast, very few recent studies formally incorporate and test the influence of the textual content of user-generated reviews [15,16]. Researchers [11] have analyzed online user-generated reviews for the hotel industry (eg, TripAdvisor) to better measure hotel service quality and performance. They have attempted to create a more comprehensive view of online user-created content by considering both quantitative aspects and textual context from multidimensional perspectives by applying text classification and sentiment analysis techniques. Further, there is a hospital service quality version of sentiment analysis that uses Twitter data. For instance, Chou et al [17] found that the “Mayo Clinic” was rated as negative, but an examination of tweets mentioning the clinic showed a mix of sentiments. Some that were rated negative (78%) were associated with concern about someone in surgery or with having an illness rather than with care or quality.

Although various types of social media can be used to obtain health-related information, online health communities are the most pertinent for several reasons:

1. We can focus on social interaction between those who question and those who answer. In online health communities, patients and their caregivers can share their experiences and exchange hospital and medical information. In particular, emotional impressions and qualitative evaluations of hospitals offered by community members are crucial points for patients and caregivers when deciding which hospital to choose.
2. Convenience and anonymity are important reasons why mothers use the Internet [18]. They expect to find health-related and hospital-related information easily and quickly, and they are typically not embarrassed to ask questions of online health professionals or communicate with online members about their personal conditions and experiences [19].
3. Online health communities enable patients to take a more active role when choosing an appropriate hospital through the use of social support. They can consider multiple aspects such as “service”, “professionalism”, “process”, “environment”, “impression”, and “popularity”.

Therefore, we selected user-generated messages from online health communities for our experimental data.

We aimed to reveal the details of hospital reputation by not only conducting sentiment analysis but also analyzing multidimensional service quality factors inside users' written messages. Our proposed approach does not only rely on the number of “likes” or “recommendations” but finds out which hospital service quality factors are really associated with the recommendations of online health communities. In summary, our contributions are threefold:

1. We define the social media–based hospital service quality factors model consisting of six service quality factors: “service”, “professionalism”, “process”, “environment”, “impression”, and “popularity”, based on observation of the content in online health communities.
2. We propose a novel approach to detect social media–based hospital service quality factors using text mining techniques. Our key aim in this task was to automatically identify these factors in online health communities effectively. To this end, we implemented a set of text mining modules to extract potential clues to those service quality factors and to classify them into two recommendation types. Detailed steps include focused crawling, preprocessing, dictionary-based hospital name extraction, filtering, detecting quality factors for hospitals, and recommendation classification.
3. We provide the analysis results that can be visualized as patterns of major behavior change factors (for hospitals that were negatively assessed by the general public) by analyzing the six main cities in Korea. The analysis shows the effectiveness of our proposed methods to reveal the potential of collective intelligence based on social media–based hospital service quality factors.

Methods

Data

To investigate the opinions about the quality factors of hospitals in South Korea, particularly the quality of pediatric hospitals, we collected 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 hospitals, sharing their thoughts, ideas, and experiences. Considering that Naver [20] and Daum [21] operate the dominant local online communities for parents and caregivers, we decided to analyze the content of these two portals.

Our aim was to collect user-generated content that could contain users’ experiences with and opinions about pediatric hospitals from online health communities. The structure of the content can be divided into three levels: threads, messages, and sentences. Threads often contain several messages, whereas messages are short and often comprise only a few sentences or sentence fragments. Figure 1 is a snapshot of the hospital recommendation content of online communities, where (2) is the thread title, (4) indicates the details of the thread, and (5) contains the messages of the thread.

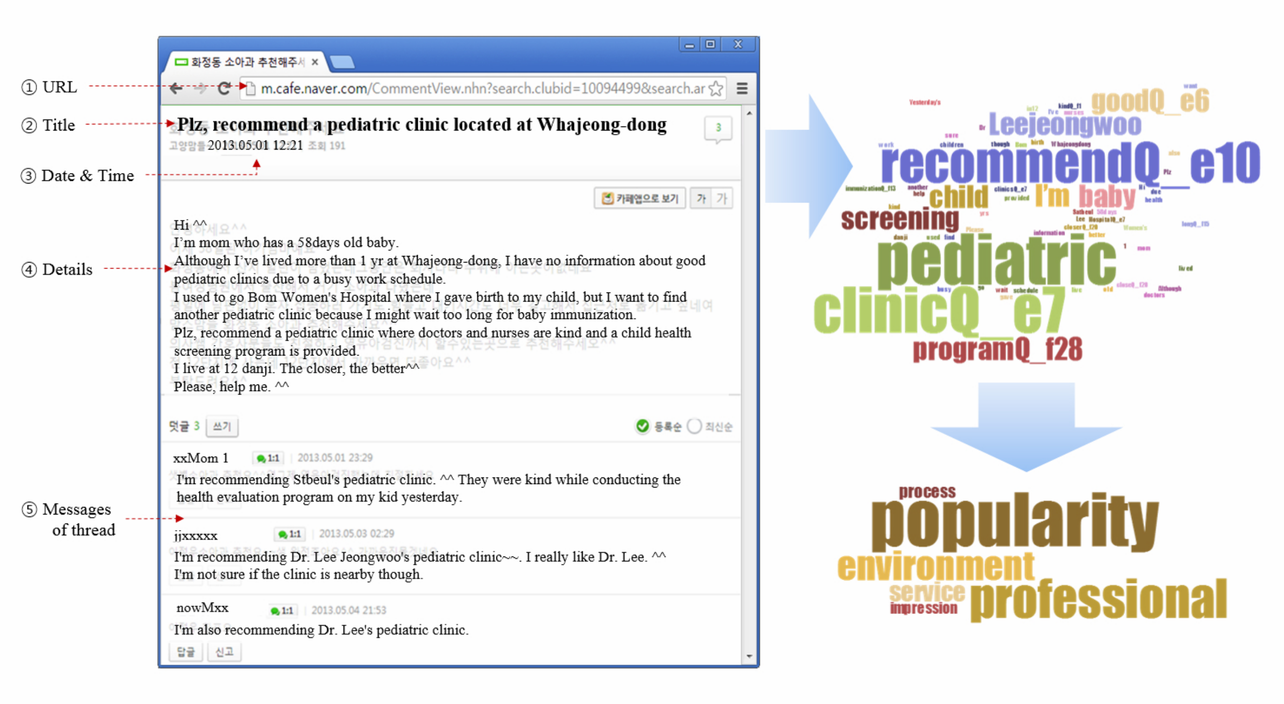
Table 1. Data statistics (from Naver and Daum Web portals).

District	Threads, n	Messages, n	Messages per thread, n	Messages containing quality factors, n
Seoul	10,832	54,392	5.02	12,421
Daegu	8072	47,419	5.87	4240
Busan	5965	28,910	4.85	9509
Daejeon	3952	22,475	5.69	2358
Incheon	775	5184	6.69	1525
Gwangju	2826	15,368	5.44	2012
Total	32,422	173,748		32,065 (18.45%)
Average	5403.66	28,958	5.59	

Table 1 shows statistics on the content from the selected parents/caregiver online communities from Naver and Daum between April 2007 and May 2013. We obtained 32,422 threads and 173,748 messages. On an average, there were 5.59 pediatric hospital-related discussions per thread among the selected regions’ community members. More specifically, 18.45% out of the total messages contain clue expressions for hospital quality factors.

It is worth noting that over the sample period, there were increases in both Internet usage rate in South Korea and the content in online health communities on the two portal sites. As of 2013, the number of Internet users in South Korea was 40,080,000, (ie, 82.09% of the population) [22]. About 50% of Internet users accessed social media for personal use, and more than 70% of users have used an online community between 2010 and 2012. In the case of user-related content in online health communities, the number of threads has rapidly increased since its launch in 2003 but the number seems to have stabilized since 2010. In Figure 2, we present both these trends; our proposed method will properly address such changes in volume (the number of threads for 2013 was extrapolated using data available up to May 2013).

Figure 1. An example of online community and extracted quality factors for hospitals.



Plz, recommend a pediatric clinic located at Whajeong-dong

2013.05.01 12:21 조회 191

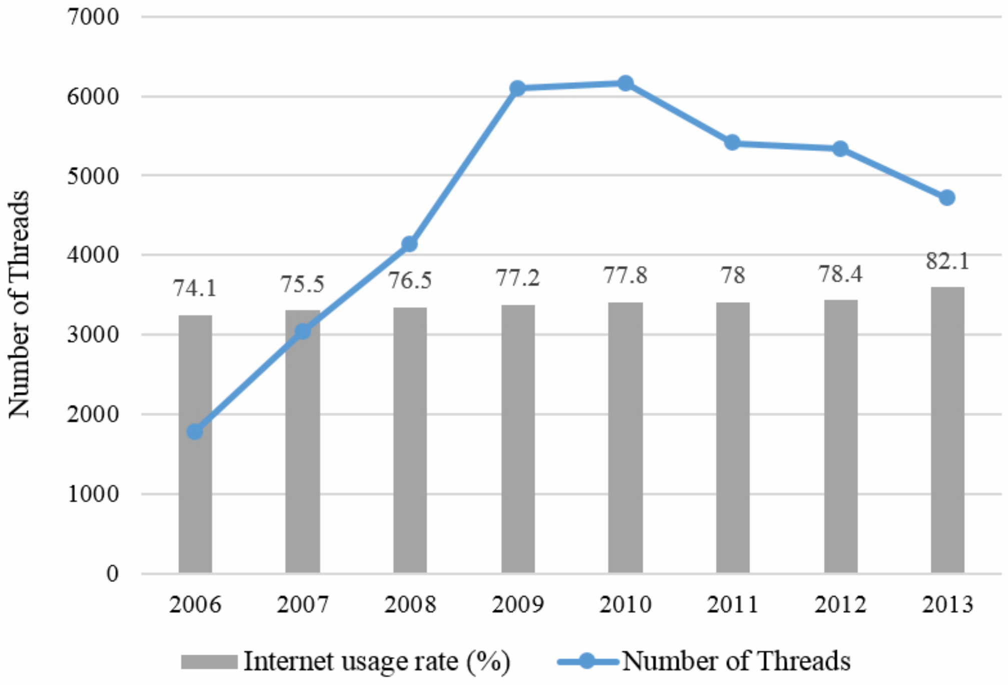
Hi ^^
I'm mom who has a 58days old baby.
Although I've lived more than 1 yr at Whajeong-dong, I have no information about good pediatric clinics due to a busy work schedule.
I used to go Bom Women's Hospital where I gave birth to my child, but I want to find another pediatric clinic because I might wait too long for baby immunization.
Plz, recommend a pediatric clinic where doctors and nurses are kind and a child health screening program is provided.
I live at 12 danji. The closer, the better^^
Please, help me. ^^

xxMom 1 1:1 2013.05.01 23:29
I'm recommending Stbeul's pediatric clinic. ^^ They were kind while conducting the health evaluation program on my kid yesterday.

jjxxxxx 1:1 2013.05.03 02:29
I'm recommending Dr. Lee Jeongwoo's pediatric clinic~~. I really like Dr. Lee. ^^
I'm not sure if the clinic is nearby though.

nowMxx 1:1 2013.05.04 21:53
I'm also recommending Dr. Lee's pediatric clinic.

Figure 2. Internet usage rate and number of threads.



Derivation of Social Media–Based Key Quality Factors for Hospitals

To efficiently manage social media data in terms of hospital service quality measurement, we defined social media-based hospital service quality factors by referring to previous quality models. However, it is also worth noting that there exist several important research studies [23-29] on service quality models for hospitals. In terms of measuring patient satisfaction, the SERVQUAL model [23], which consists of five dimensions: (1) responsiveness, (2) reliability, (3) assurance, (4) tangibles, and (5) empathy, is the most widely used tool [24]. The SERVQUAL model helps obtain customer ratings of perception and expectation on an ordinal scale. Due to its innate incompleteness, several approaches employ a modified SERVQUAL approach. For example, [25] added two quality dimensions, "caring" and "patient outcomes", to SERVQUAL, and [26] added "core medical outcomes" and "professionalism/skill/competence".

In terms of consumer perspectives, five quality factors for health care providers can be defined: (1) warmth, caring, and concern,

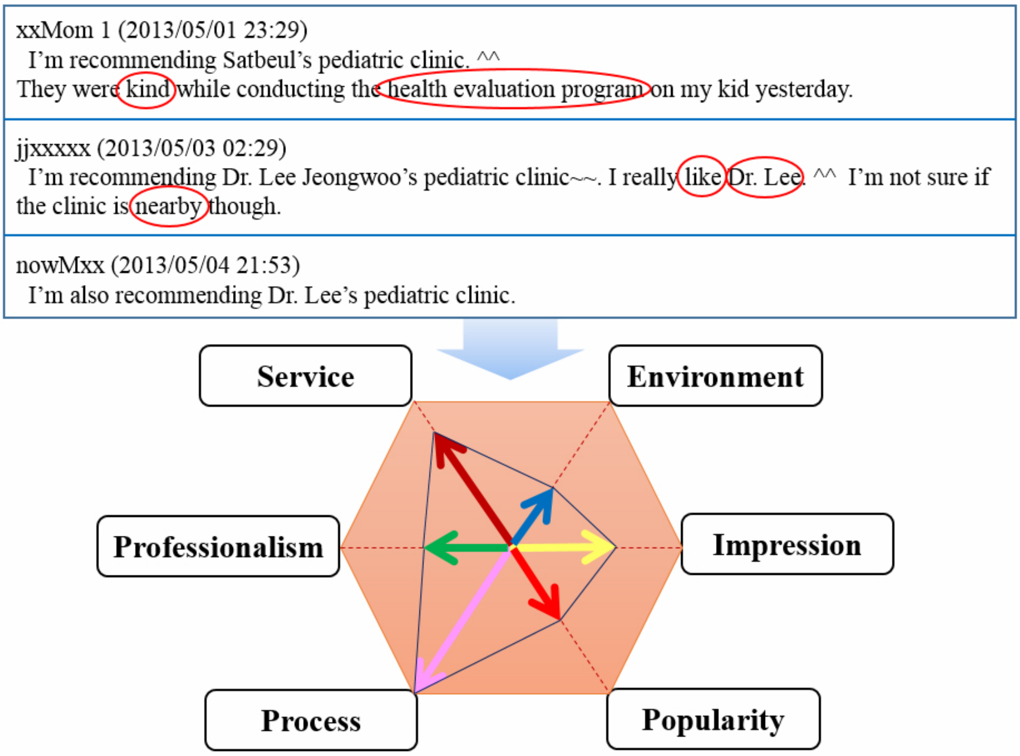
(2) medical staff, (3) technology and equipment, (4) specialization and scope of services available, and (5) outcome [27].

For use in the hospital accreditation process, the Joint Commission [28] defined nine quality dimensions for hospitals: (1) efficacy, (2) appropriateness, (3) efficiency, (4) respect and caring, (5) safety, (6) continuity, (7) effectiveness, (8) timeliness, and (9) availability. The Joint Commission model is more comprehensive because it encompasses SERVQUAL and the five quality factors for health care providers.

In an effort to develop a valid and reliable instrument for hospital management's strategic and operational decision-making, the Key Quality Characteristics Assessment for Hospitals model was developed [29], which consists of eight factors: (1) respect and caring, (2) effectiveness and continuity, (3) appropriateness, (4) information, (5) efficiency, (6) meals, (7) first impressions, and (8) staff diversity. The model used inputs from both health care providers and consumers. However, the existing service quality measurement models are not appropriate for interpreting social media data in terms of a hospital's service quality because users write their experiences without predefined formats or principles.

To this end, we manually listed all the keywords that appeared more than five times in our corpus. Then we selected only terms related to services quality, for example terms regarding the status, impression, and treatments of hospitals. We finally resolved the model for social media–based hospital service quality factors to consist of six quality factors (ie, "service", "professionalism", "process", "environment", "impression", and "popularity"), as shown in Figure 3. Each service quality factor has an average of 1000 keywords as features that can be used for analyzing the content of user-written messages. The six service quality factors have their own subcategories and subordinating items, as listed in Table 2. Each quality factor has more than one subcategory, numbered from f1–f29 and e1–e10, according to their functional or emotional characteristic, respectively.

Figure 3. Social media-based key quality factors for hospitals.



xxMom 1 (2013/05/01 23:29)
I'm recommending Satbeul's pediatric clinic. ^^
They were kind while conducting the health evaluation program on my kid yesterday.

jjxxxxx (2013/05/03 02:29)
I'm recommending Dr. Lee Jeongwoo's pediatric clinic~~. I really like Dr. Lee. ^^ I'm not sure if the clinic is nearby though.

nowMxx (2013/05/04 21:53)
I'm also recommending Dr. Lee's pediatric clinic.

Table 2. Detailed categorizations and subordinate items of social media–based hospital service quality factors.

Quality factor	Detailed quality factor	Description
Service	Kindness (f1/f2)	Kindness, courtesy related to doctor/nurse/hospital staff manners
		Diagnose and explain in easy words/rough detail
	Faithfulness (f3/f4)	Response to a request faithfully/carefully
Professionalism	Professionalism (f5/f6/f7)	Technical knowledge, skills
		Special medical courses, professional
		Professor, director (rank)
	Skill (f8/f9/f10)	Medical procedures/skills, experience
		Side effects, complications, medical malpractice, safety
		Treatment effects, speed
	Treatment (f11/f12/f13/f14)	Hospitalization, outpatient, inpatient
		Accuracy
		Antibiotics, injections, prescription drugs (powder, liquid medicine, cold medicine)
		Diagnosis, treatment, prescription, cure, appropriate (or over-) treatment
Process	Speed (f15/f16)	(Short/long) waiting time, dose standby, receiving the relevant treatment
		Responding immediately to changes, ability to cope with emergencies
	Cost (f17)	Low (or reasonable)/expensive medical costs
	System (f18/f19)	Efficient business processes (reception and express services), questionnaires, (basic/optional) medical care
Environment	Convenience (f20/f21/f22/f23)	Transportation, distance (from residence), parking facilities
		Reservations
		Waiting room, convenient facilities
		Office hours, dates, evening hours (weekends, Sundays, and late hours), closed hours
	Sanitation (f24/f25)	Cleaning, management
		Disposable products, sanitary ware
	Facility/ Equipment (f26/f27/f28/f29)	Hospitalization
		CT, MRI, equipment, tools, operating room, doctor’s office
		Surgery, physical therapy, various tests, therapies, health screenings, hospital rounds
		Hospital size
		Types of hospital (public health center, university hospital, hospitals, clinics, and private hospitals)
Impression (e1/e2/e3/e4/e5)		Image of the hospital, atmosphere
		Reliability, favorite hospitals, physician
		Signs, new doctors, or hospital encountered for the first time
		Tired of existing hospital
		Impression of the doctor, doctor’s information (ie, mood, personality, gender, age)
Popularity (e6/e7/e8/e9/e10)		Rumor, tradition, reputation
		Hospital name
		Media, advertising
		Well-known doctor (doctor’s name)
		Anyone who knows the hospital

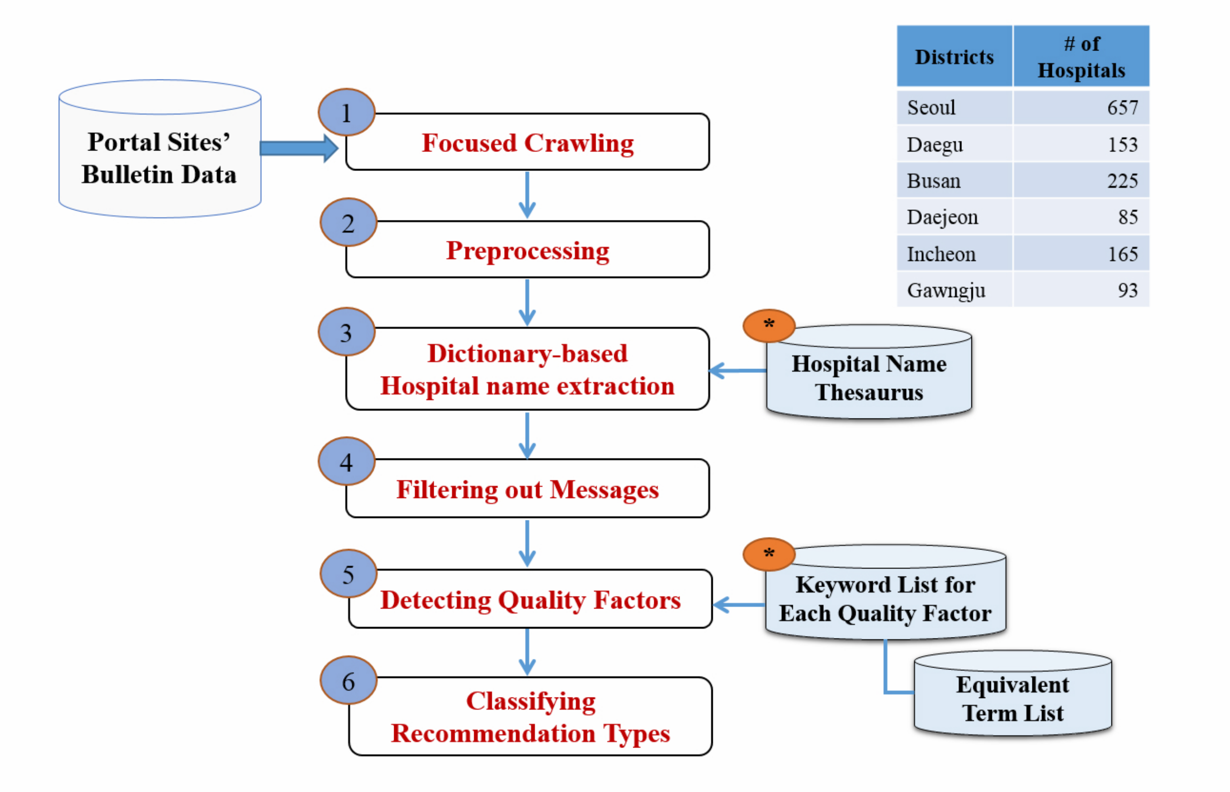
Text Mining for Hospital Quality Factors Extraction and Recommendation Classification

Overview

We implemented a set of text mining modules for our social media–based hospital service quality factors model-based extraction and classification. The process consists of six steps as shown in Figure 4: data preparation through crawling, data preprocessing, hospital name extraction, message filtering, quality factor detection, and recommendation type classification.

After data aggregation via focused crawling, the HTML elements are removed. For every message, the next four natural language processing steps are applied to detect quality factors in the messages and to compute the hospitals' recommendation type based on them. Although our approach can be applied to various types of domains in online health communities, we targeted messages for analysis that explicitly mentioned pediatric hospitals/clinics and discussed the quality factors related to those hospitals/clinics.

Figure 4. Steps for detecting quality factors and recommendation classification.



Districts	# of Hospitals
Seoul	657
Daegu	153
Busan	225
Daejeon	85
Incheon	165
Gawngju	93

Step 1. Focused Crawling

The two Web portals, Naver and Daum, operate various kinds of online communities arranged by region and subject. For example, the Naver portal has approximately two million regional communities. We note that each region includes at least one parent community with active members, high popularity, and vigorous online activity. We opted for these local online communities as our text mining information sources. Each online community provides a search function that allows us to identify particular webpages with "pediatric" and "recommendation" keywords. The webpages from the selected online parent communities were crawled and stored in HTML format.

Step 2. Preprocessing

HTML tags and non-textual information such as images, JavaScript codes, and advertisements were deleted from the extracted files. For effective handling of hypertext markup language (HTML) content, we used the Beautiful Soup library, designed in the Python programming language [30].

Step 3. Dictionary-Based Hospital Name Extraction

Often, when people mention hospitals in a social context in Korean, they use a range of hospital names, mostly expressed through acronyms or abbreviations. To counter this problem, we built a hospital name dictionary covering the pediatric hospitals in the six big cities (ie, Seoul, Daegu, Busan, Daejeon, Gwangju, and Incheon) in South Korea. The hospital names were extracted from the Health Insurance Review and Assessment Service website [31]. This dictionary is a value (expression) mapping table that adds acronyms or abbreviations for every hospital name observed in user-written sentences in the online communities. Using this dictionary, our text mining module performs a specially designed stepwise expression normalization procedure, based on the textual edit-distance similarity [32] 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. Filtering Out Messages

Once the canonical hospital names were extracted from the messages, we filtered out messages in which hospital names were not mentioned or which were ambiguous, so as to consider only messages that mentioned hospital names clearly. In addition, we selected candidate messages that recommended or in some other way mentioned hospitals.

Step 5. Detecting Quality Factors

This is the main step for identifying textual representations of quality factors for hospitals. By utilizing the keyword lists (ie, a dictionary of quality factors), we detected and extracted the service quality factors from each message. The techniques used in this task are derived from dictionary-based named entity recognition [33]. If service quality factors are identified, we count the number of occurrences for further analysis. In addition, to deal with known difficulties of Korean language processing, we used an equivalent term list derived from high frequency terms observed in our data collection.

Step 6. Classifying Recommendation Types

Based on the expressions of identified quality factors, this step classifies the recommendation type of each message, that is, whether a hospital is recommended or not. As a simple implementation of keyword-based sentiment analysis, it counts the number of sentimental keywords for and against recommendation [9]. To consider possible changes of type within sentences in a single message, we use a modified version of mood flow analysis [34]. As results, the classified types are positive (ie, recommended), neutral, bilateral, and negative (ie, not recommended). If we simply summarize the classification results, 35.86% (47,046/131,191) are hospital recommendations where 131,191 is the total number of quality factors considered as one of four classified types, and 47,046 is the number of quality factors showing positiveness. It is to be noted that a single message can have more than one qualify factor. However, we reduced the classification categories to two types (ie, positive or negative) by viewing it as a binary classification problem; misclassification of outcome type is a serious concern, particularly when two different polarities coexist in the same message.

Manual Evaluations

Before analyzing in depth the results of Steps 5 and 6 in various ways, we performed manual evaluations on the extraction and classification results for things such as hospital names, quality factors, and recommendation types. For the evaluation of hospital name extraction, we performed human tagging of a random sample of 9450 messages, which amounted to 5.4% of the total messages. In addition, for the evaluation of quality factor detection and recommendation type classification, we selected the five most frequently occurring hospitals/clinics from the six regions and manually checked their subordinating quality factors and recommendation type to check

if their extractions/classifications were correct. The evaluation results are summarized in Table 3.

Table 3. Evaluation results (reported as percentages).

Region	Evaluation category						
	Hospital name extraction			Quality factors detection			Recommendation type classification
	Precision	Recall	F1	Precision	Recall	F1	Precision
Seoul	68	83	75	93	95	94	76
Daejeon	91	72	80	96	91	93	82
Daegu	86	65	74	97	80	88	70
Gwangju	91	66	77	96	83	89	64
Incheon	79	77	78	88	93	90	61
Busan	87	73	79	98	82	89	67
Average	84	73	78	95	87	91	78

Hospital Name Extraction

We achieved a recall (the proportion of hospital names in the original text that were extracted correctly) of 73% and precision (the proportion of extracted hospital names that were correct) of 84%, giving F1 = 2 × (precision) × (recall)/(precision + recall) = 77.7%. The measure F1 is the harmonic mean of precision and recall. Although accuracy measures of 80%-90% have often been achieved in named entity extraction in English, there are very few cases of such high accuracy in informal texts written in Korean due to various types of acronyms, misspellings, and frequently incorrect spacing between words. These factors are currently performance bottlenecks together with the fact that Korean is an agglutinative language.

Detecting Quality Factors

Quality factor detection is a kind of keyword matching that checks if a given word exists in one of the categories of the six quality factors. Its performance is slightly over 90% due to errors in word spacing and homonym problems of the Korean language. The currently achieved F1 score is considerably satisfactory. However, the detection performance could be enhanced further by considering compound nouns and more sophisticated processing of erroneously written sentences for clearer word spacing.

Recommendation Classification Using Sentiment Analysis

Our recommendation classification, a type of sentiment analysis, is a challenging task in that it should track polarity transitions within a message. In spite of handling polarity transitions within a message by referring to [34], the verified classification performance is 77.8% on average. This indicates that sentiment analysis in Korean still has room for further improvement. Nevertheless, our quality factor detection and sentiment analysis provides the option to facilitate the in-depth analysis of messages by region.

Results

Overview

By applying the text mining techniques introduced earlier to online health community messages, we obtained a collective intelligence driven by our social media–based hospital service quality factors model. As a first step to investigate the potential of this knowledge, we analyze the occurrence patterns of preferred quality factors among different big cities in South Korea.

Overall Social Media–Based Hospital Service Quality Factors Distributions by Region and Time

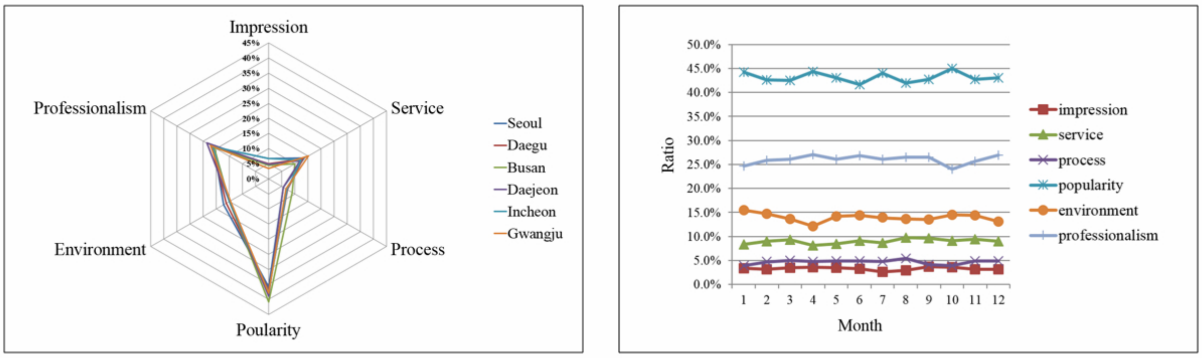
Previously, labor-intensive surveys have been used for hospital service quality evaluation. In contrast with such costly manual surveys, we can easily access multidimensional analysis results from the occurrence ratios of keywords included in one of the six quality factor categories. The occurrence ratios of each quality factor signify caregivers' preferences when they select pediatric hospitals/clinics.

Panel (a) of Figure 5 gives a high-level view of the six key quality factor distributions for the six largest cities (ie, Seoul, Daegu, Busan, Daejeon, Incheon, and Gwangju) in South Korea. The plots in panel (a) of Figure 5 indicate the occupying ratios of each quality factor for each region. Although there are small variations among the different six quality factors, two distinctive factors (popularity and professionalism) occur most frequently.

Such trends are maintained every month, as shown in panel (b) of Figure 5, where the raw frequencies of quality factors are averaged over all regions. In contrast, when we observe regional data, the mean relative frequency ratios of their constituting factors differ somewhat by region and season. Panels (c) and (d) of Figure 5 show the mean relative frequency ratios per month during 2006-2013 for two selected cities, Daejeon and Gwangju, respectively. The numbers in panels (b)-(d) of Figure 5 are given as percentages of occurrence.

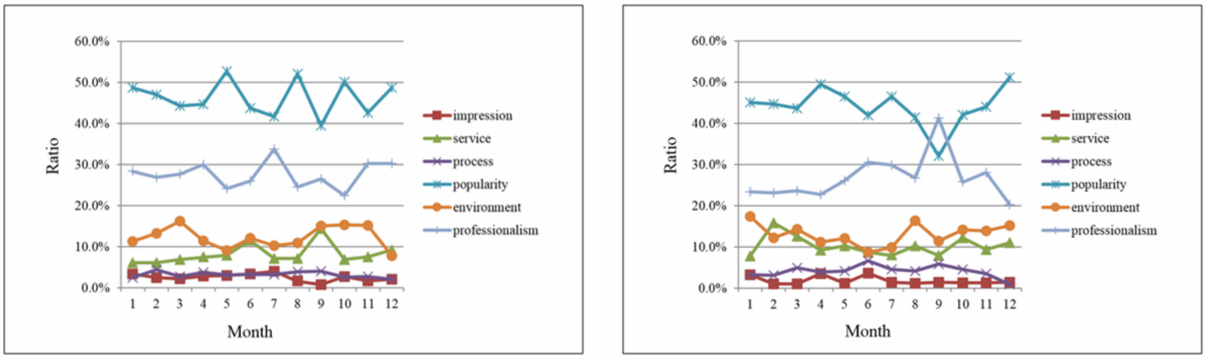
We assume that the analysis results driven by social media–based hospital service quality factors form a kind of collective intelligence that represents caregiver beliefs regarding the quality of local hospitals and can be effectively used to discover hospital recommendations within online health communities. Our analysis provides insight into regional and seasonal trends in quality factors that can be obtained from observing pertinent keywords/terms.

Figure 5. An example of overall quality factor distributions.



(a) Radiation chart of six big cities

(b) Average of six cities



(c) Daejeon

(d) Gwangju

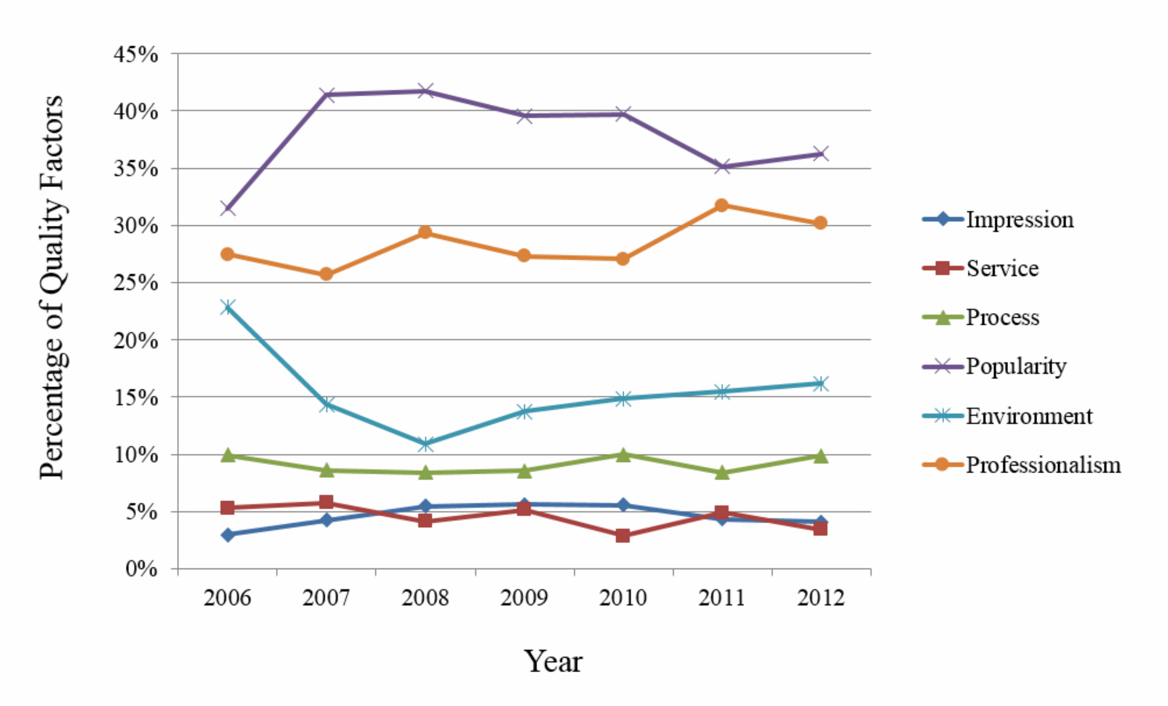
Case Study of Negative Attitudes

Overview

Based on the firm belief that there exist clear differences in the distributions of social media–based hospital service quality factors according to recommendation types, we disaggregated the negative parts from the previous results to reveal hidden regional relationships between the factors. For this, we used sentiment analysis that decomposes user-written messages into two types (ie, recommended and not recommended).

A large number of users may have positive or neutral attitudes about their hospital experiences, even though one certain user had a negative attitude. Based on our observations, on average, 8% of the samples were identified as negative cases if we include conflicting cases (where positive and negative sentiment coexist in a single message) as negative ones. Figure 6 shows the percentages of negative comments on social media–based hospital service quality factors per year during 2006-2013 for the six major cities of South Korea. We could say that it shows the domestic characteristics of a negative attitude towards pediatric hospitals.

Figure 6. Negative attitude trends by year (for six cities).



Influenza Pandemic in 2009

We draw this conclusion from the influenza pandemic in 2009. Not only South Korea but the world suffered from a new influenza A in 2009 [35]. The most severe period for South Korea was approximately from October to November of that year; the effect of the pandemic slowly reduced after December 2009. We verified 78% of influenza-related posts collected from online health communities that appeared in 2009. The details of posts mostly consisted of questions and answers, about which hospital had vaccines to spare, or how long it would take to see a doctor and get medicine prescribed. In reality, the supply of influenza vaccine was considerably insufficient, and it is clear that most patients had trouble finding hospitals to immunize their children. As people streamed into local hospitals, they posted complaints about long waiting times. In other words, health care quality received a poor grade for that period.

Figure 6 shows caregivers’ strong negative responses in 2009. Since there was also an increase in the volume of user-generated content in 2009 as seen in Figure 2, we used the percentage of quality factors instead of the mere frequency. In fact, although the occurrence of negative content regarding all six quality factors increased, we observed that the relative shares of “popularity” and “professionalism”, the top and second major factors, significantly decreased in 2009. Interestingly, we observed increases in the relative shares of “process” and “environment” in 2009, the sum of which also accounts for a significant proportion of frequencies. This implies that the experience of parents and caregivers in pediatric hospitals was often poor and their dissatisfaction increased, especially with regard to inefficient treatment process such as long waiting times or poor experience with reservation. To observe social media–based hospital service quality factors more closely, we looked at the interrelated quality factors in terms of the long waiting times at hospitals, such as process and environment, as shown in Table 4.

For the “process” factor, the top three most frequently occurring items were “long term”, “right now”, and “hospitalized” at 73%, 11%, and 10%, respectively. For the “environment” factor, the different types of keywords that characterized each corresponding factor revealed the public’s discontent and complaints about hospital services.

Table 4. In-depth analysis of negative attitude in 2009.

Process		Environment	
Items	Share, %	Items	Share, %
Long term	73	Reservation	48
Right now	11	Shot	13
Hospitalized	10	Examination	9
Basic	5	Health care center	7
Other	1	Other	23

Emergency Management in Local Cities

Patients and caregivers occasionally visit general hospitals for emergency care. As can be seen in Figure 7, we observe that the “emergency”-related threads in the Daegu region increased steadily and continuously over several years until 2010. Presuming that they represent an emergent event posted in online communities, we note that approximately 5% of the threads talk about whether particular hospitals provide prompt and satisfactory emergency medical services.

In many cases, general hospitals are expected to have sufficient capacity to deal with an emergency. For example, a person who has an injured child desperately seeks hospitals open at midnight. Such people may think that the first general hospital able to provide care has competent medical staff and the necessary emergency care equipment. However, the reality in local cities is often not satisfactory. Table 5 shows that emergency-related terms often co-occur with five local general hospital names in many negatively opinionated threads (and are classified as not recommended). In this regard, we become familiar with which quality factors need to improve by analyzing social media–based hospital service quality factors. For example, “process” and “environment” factors are often detected in threads related to emergency.

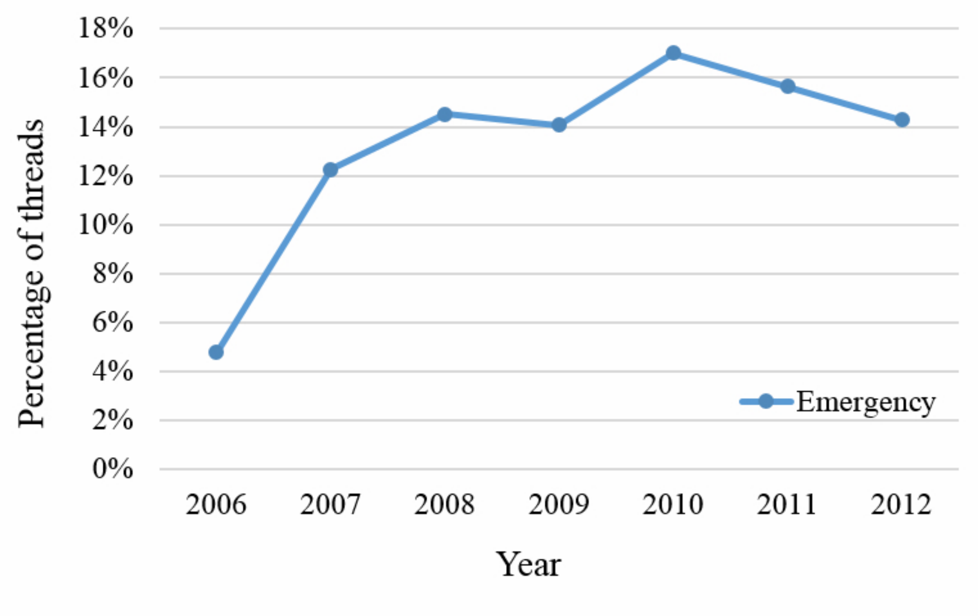
Table 5. Co-occurrence ratios in negatively opinionated threads (Daegu region).

Co-occurrence pair	Percentage
Emergency & Hospital D1	22
Emergency & Hospital D2	16
Emergency & Hospital D3	7
Emergency & Hospital D4	31
Emergency & Hospital D5	8

One of the key reasons for this phenomenon is the small number of pediatric registrars in local general hospitals. The problem is endemic: applicants to the pediatric residency programs of local general hospitals in Daegu are lacking. Only the top two general hospitals met the required enrollment numbers in 2011. The problem can lead to poor quality of medical services, including first aid and emergency treatment. Some journalists have pointed out the need to improve hospital capacity, capability, and preparedness in the event of an emergency. The other reason is that hospital information in the local area is not well circulated. People often could not find adequate hospitals in an emergency, so they posted questions about which hospitals were available.

If the key quality factors for hospitals are detected and analyzed automatically and periodically, we can determine what kind of medical services are necessary to contain a highly contagious epidemic. Health authorities can implement a policy that could help patients based on immediate and tangible qualitative observations. From the point of view of a patient, they could discover which negative quality factors were prevalent among nearby hospitals, and hence find better medical services in other areas. Moreover, the analyzed patterns are meaningful advisory information for government policy makers because they explain the claims of caregivers who go to pediatric hospitals. Such information is helpful for both national and local health policy makers.

Figure 7. Percentage of threads that mention “Emergency” in the Daegu region.



Discussion

Principal Findings

With the growing popularity of Web 2.0, a number of caregivers who have children are participating in reviewing, recommending, and exchanging opinions about pediatric hospitals. In this regard, a thorough understanding of the qualitative measures of a hospital could be important for selecting an appropriate pediatric hospital/clinic. Therefore, our contributions are threefold: (1) we define six types of quality factors feasible for social media–based hospital service quality analysis, (2) we propose a text mining approach to extract several hospital quality factors and to classify the messages into recommendation types (recommended or not recommended), and (3) we provide an in-depth analysis that can be used for many shareholders including caregivers, hospital staff, and government officials of health care policy.

Limitations

Despite the value of the new analysis generated by this work, this study has several limitations in terms of data coverage. Social media, like other tools that assess public opinion, lends itself to selection bias, and it is difficult to assess whether this study population represents a fair sample of the general population [36]. In addition, owing to the limited information (accounting for regional tendencies), more in-depth analysis as to why different fluctuation patterns occur is required. However, it goes beyond the scope of this study, as the reasons are highly correlated with demographic characteristics.

Problems such as the inherent anonymity of users and the inability to authenticate content can be slightly mitigated because, in contrast to Twitter, messages from online health communities contain more detailed information originating from real personal experiences at hospitals. As the coverage of recommendation of pediatric hospitals/clinics in social media is not without its pitfalls, the findings highlight the potential for social media to perpetuate and perhaps even promote sources of information of unknown quality.

This study is based on our firm belief that social media can be used to gauge how patients and caregivers respond to a particular hospital. However, it has been previously found that social media also has risks. Broader scope and easily accessible platforms can lead to the wide dissemination of misinformation. We may overestimate the effect of our quality factor metrics because of the misinterpretation or over-simplification of findings from other socio-scientific studies. Nevertheless, our experimental analysis shows the great potential for social media–based key quality factor analysis for hospitals.

Next Steps

A natural next step would be to validate the results of our analysis by comparing hospital service quality as measured through the Internet user-related data with traditional quality measures as well as patient satisfaction surveys. Previously in the literature, Greaves et al [5] did sentiment analysis on online free-text comments about hospitals and compared the results with the paper-based national inpatient survey results for all 161 acute hospitals with adult services in the United Kingdom. They found that the online comments are moderately associated with the survey and thus online comments can be used to assess patients’ opinion about hospital performance. While they compared hospital ranks, they did not look into the detailed categories of hospital service quality factors that we attempted in this study. In fact, to the best of our knowledge, a patient survey on hospital service quality factors has never been conducted at the national level. Currently in Korea, there exist several government-run websites that provide information on hospitals, for example the Ministry of Health and Welfare [37], Statistical Information Service of Health and Welfare [38], and Health Insurance Review and Assessment Service [31]. However, none of these accreditation bodies provide hospital evaluation information at the patient's satisfaction level, especially in terms of the six service quality factors used in this study. We searched and found academic studies detailing sentiment analysis of social network service mentioning about hospitals in Korea utilizing the SERVEQUAL model [39,40]. However, Song [39] analyzed social media comments about three hospitals only for 3 months. Although Kang and Song [40] compared their sentiment analysis results with survey data and verified that social media results corresponded with survey results, their analysis was confined to the top five hospitals in Korea. In fact, there have been few studies verifying social media content about hospital quality to date, mainly due to the lack of hospital quality surveys conducted at the national level. If a survey of hospital service quality will be conducted at the national level, it would be important and interesting to validate how well our social media-based analysis reflects the actual quality of hospitals.

Although we leave such validation work to be conducted in future studies, real-time information extracted from online sites via our proposed methodology can be still helpful to government authorities as well as information-seeking patients at both local and national levels. Recently, information from social media has been analyzed and used at the national level in the United Kingdom. In particular, the UK government recommends the use of social media and sentiment analysis for rapidly measuring hospital performance and poor care [41] and also runs the website National Health Service (NHS) Choices [42], where people can rate the quality of care received by their family practice.

Conclusions

This work and future studies in this field are critical to inform both research and health care communities about the current status of hospital service quality and to highlight the need for careful, evidence-based analysis of user-preferable quality factors and recommendations based on them. To further explain the possible causes of hospital selection patterns, we plan to analyze the results in conjunction with demographic data.

Conflicts of Interest

None declared.

References

1. Lu Y, Zhang P, Liu J, Li J, Deng S. Health-related hot topic detection in online communities using text clustering. PLoS One 2013;8(2):e56221 [FREE Full text] [doi: 10.1371/journal.pone.0056221] [Medline: 23457530]
2. 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]
3. Emmert M, Meier F, Heider AK, Dürr C, Sander U. What do patients say about their physicians? an analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
4. 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;14(1):e17 [FREE Full text] [doi: 10.2196/jmir.1868] [Medline: 22569642]
5. 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]
6. 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 [FREE Full text] [doi: 10.2196/jmir.3078] [Medline: 24496094]
7. Goldstein E, Farquhar M, Crofton C, Darby C, Garfinkel S. Measuring hospital care from the patients' perspective: an overview of the CAHPS Hospital Survey development process. Health Serv Res 2005 Dec;40(6 Pt 2):1977-1995 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00477.x] [Medline: 16316434]
8. Hiidenhovi H, Nojonen K, Laippala P. Measurement of outpatients' views of service quality in a Finnish university hospital. J Adv Nurs 2002 Apr;38(1):59-67. [Medline: 11895531]
9. Pang B, Lee L. Opinion Mining and Sentiment Analysis. FNT in Information Retrieval 2008;2(1–2):1-135. [doi: 10.1561/1500000011]
10. 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]
11. Guy I, Avraham U, Carmel D, Ur S, Jacovi M, Ronen I. Mining expertise and interests from social media. In: Proceedings of the 22nd International Conference on World Wide Web.: International World Wide Web Conferences Steering Committee; 2013 Presented at: WWW '13 - 22nd International World Wide Web Conference; May 13-17, 2013; Rio de Janeiro, Brazil p. 515-526.
12. Archak N, Ghose A, Ipeirotis PG. Deriving the Pricing Power of Product Features by Mining Consumer Reviews. Management Science 2011 Aug;57(8):1485-1509. [doi: 10.1287/mnsc.1110.1370]
13. Chevalier JA, Mayzlin D. The Effect of Word of Mouth on Sales: Online Book Reviews. Journal of Marketing Research 2006 Aug;43(3):345-354. [doi: 10.1509/jmkr.43.3.345]

14. Duan W, Gu B, Whinston AB. Do online reviews matter? — An empirical investigation of panel data. Decision Support Systems 2008 Nov;45(4):1007-1016. [doi: 10.1016/j.dss.2008.04.001]
15. Ghose A, Ipeirotis PG. Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. IEEE Trans. Knowl. Data Eng 2011 Oct;23(10):1498-1512. [doi: 10.1109/TKDE.2010.188]
16. Netzer O, Feldman R, Goldenberg J, Fresko M. Mine Your Own Business: Market-Structure Surveillance Through Text Mining. Marketing Science 2012 May;31(3):521-543. [doi: 10.1287/mksc.1120.0713]
17. Chou WYS, Hunt Y, Folkers A, Augustson E. Cancer survivorship in the age of YouTube and social media: a narrative analysis. J Med Internet Res 2011;13(1):e7 [FREE Full text] [doi: 10.2196/jmir.1569] [Medline: 21247864]
18. Umefjord G, Petersson G, Hamberg K. Reasons for consulting a doctor on the Internet: Web survey of users of an Ask the Doctor service. J Med Internet Res 2003 Oct 22;5(4):e26 [FREE Full text] [doi: 10.2196/jmir.5.4.e26] [Medline: 14713654]
19. Anderson JG, Rainey MR, Eysenbach G. The impact of CyberHealthcare on the physician-patient relationship. J Med Syst 2003 Feb;27(1):67-84. [Medline: 12617199]
20. NAVER. URL: http://www.naver.com/ [accessed 2013-10-28] [WebCite Cache ID 6Ki54BAEM]
21. DAUM. URL: http://www.daum.net/ [accessed 2013-10-28] [WebCite Cache ID 6Ki7w7xVe]
22. Korea Internet and Security Agency (KISA). 2013 survey on the internet usage executive summary. URL: http://isis.kisa.or.kr/eng/board/?pageId=040100&bbsId=10&itemId=325 [accessed 2014-10-26] [WebCite Cache ID 6TbzdlCrk]
23. Parasuraman A, Zeithaml VA, Berry LL. A Conceptual Model of Service Quality and Its Implications for Future Research. Journal of Marketing 1985;49(4):41. [doi: 10.2307/1251430]
24. Sadiq Sohail M. Service quality in hospitals: more favourable than you might think. Managing Service Quality 2003 Jun;13(3):197-206. [doi: 10.1108/09604520310476463]
25. Bowers MR, Swan JE, Koehler WF. What attributes determine quality and satisfaction with health care delivery? Health Care Manage Rev 1994;19(4):49-55. [Medline: 7896552]
26. Ramsaran-Fowdar RR. Identifying health care quality attributes. J Health Hum Serv Adm 2005;27(4):428-443. [Medline: 16318013]
27. Coddington DC, Moore KD. Quality of care as a business strategy. Healthc Forum 1987;30(2):29-32. [Medline: 10280924]
28. Joint Commission on Accreditation of Healthcare Organizations. Comprehensive Accreditation Manual for Hospitals: The Official Handbook. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations; 1996.
29. Sower V, Duffy J, Kilbourne W, Kohers G, Jones P. The dimensions of service quality for hospitals: development and use of the KQCAH scale. Health Care Manage Rev 2001;26(2):47-59. [Medline: 11293010]
30. Leonard R. Beautiful Soup. URL: http://www.crummy.com/software/BeautifulSoup/ [accessed 2013-10-28] [WebCite Cache ID 6Ki6tNAXD]
31. Health Insurance Review and Assessment Service. URL: http://www.hira.or.kr/ [accessed 2013-10-28] [WebCite Cache ID 6Ki5geHrV]
32. Vilares M, Ribadas FJ, Vilares J. Phrase similarity through the edit distance. In: Database and expert systems applications: 15th International Conference. Berlin: Springer; 2004 Presented at: DEXA 2004; August 30-September 3, 2004; Zaragoza, Spain p. 306-317.
33. Cohen WW, Sarawagi S. Exploiting dictionaries in named entity extraction: combining semi-Markov extraction processes and data integration methods. In: Proceedings of the Tenth ACM Sigkdd International Conference on Knowledge Discovery and Data Mining.: Association for Computing Machinery (ACM); 2004 Aug Presented at: Tenth ACM Sigkdd International Conference on Knowledge Discovery and Data Mining; August 22-25, 2004; Seattle, WA p. 89-98.
34. Jung Y, Park H, Myaeng SH. A hybrid mood classification approach for blog text. In: PRICAI 2006: Trends in Artificial Intelligence: 9th Pacific Rim International Conference on Artificial Intelligence (Lecture Notes in Computer Science).: Springer; 2006 Aug Presented at: 9th Pacific Rim International Conference on Artificial Intelligence; August 7-11, 2006; Guilin, China p. 1099-1103.
35. Centers for Disease Control and Prevention. Flu URL: http://www.cdc.gov/flu/pastseasons/0910season.htm [accessed 2014-06-23] [WebCite Cache ID 6QXcz29Dh]
36. Janssens ACJW, Kraft P. Research conducted using data obtained through online communities: ethical implications of methodological limitations. PLoS Med 2012;9(10):e1001328 [FREE Full text] [doi: 10.1371/journal.pmed.1001328] [Medline: 23109913]
37. Ministry of Health and Welfare. URL: http://www.mw.go.kr/front_new/index.jsp [accessed 2014-10-26] [WebCite Cache ID 6Tc2XBphE]
38. Statistical Information Service of Health and Welfare. URL: http://stat.mw.go.kr/ [accessed 2014-10-26] [WebCite Cache ID 6Tc2g05W2]
39. Song E. A Study on the Case Analysis of Customer Reputation based on Big Data. Journal of the Korea Institute of Information and Communication Engineering 2013 Oct 31;17(10):2439-2446. [doi: 10.6109/jkiice.2013.17.10.2439]
40. Kang M, Song EJ. A Study on Customer Satisfaction Evaluation Model for Hospitals using Sentiment Analysis of Social Network Service. theebusinessstudies 2012 Nov;13(4):339-352. [doi: 10.15719/geba.13.4.201211.339]
41. Greaves F, Laverty AA, Cano DR, Moilanen K, Pulman S, Darzi A, et al. Tweets about hospital quality: a mixed methods study. BMJ Qual Saf 2014 Oct;23(10):838-846 [FREE Full text] [doi: 10.1136/bmjqs-2014-002875] [Medline: 24748372]

42. NHS Choices. URL: http://www.nhs.uk/Pages/HomePage.aspx [accessed 2014-10-26] [WebCite Cache ID 6Tc3GadKT]

Edited by G Eysenbach; submitted 26.06.14; peer-reviewed by L Ning, E Zimlichman; comments to author 09.10.14; revised version received 02.11.14; accepted 22.01.15; published 07.04.15.

Please cite as:
Jung Y, Hur C, Jung D, Kim M
Identifying Key Hospital Service Quality Factors in Online Health Communities
J Med Internet Res 2015;17(4):e90
URL: http://www.jmir.org/2015/4/e90/
doi:10.2196/jmir.3646
PMID:25855612

Original Paper

Insights Into the Impact of Online Physician Reviews on Patients' Decision Making: Randomized Experiment

Sonja Grabner-Kräuter[1], PhD; Martin KJ Waiguny[2], PhD

[1]Department of Marketing and International Management, Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria

[2]Marketing, Advertising, Retailing and Sales, Auckland University of Technology, Auckland, New Zealand

Corresponding Author:
Sonja Grabner-Kräuter, PhD
Department of Marketing and International Management
Alpen-Adria-Universität Klagenfurt
Universitätsstrasse 65-67
Klagenfurt, 9020
Austria
Phone: 43 463 2700 ext 4042
Fax: 43 463 2700 994042
Email: sonja.grabner@aau.at

Abstract

Background: Physician-rating websites combine public reporting with social networking and offer an attractive means by which users can provide feedback on their physician and obtain information about other patients' satisfaction and experiences. However, research on how users evaluate information on these portals is still scarce and only little knowledge is available about the potential influence of physician reviews on a patient's choice.

Objective: Starting from the perspective of prospective patients, this paper sets out to explore how certain characteristics of physician reviews affect the evaluation of the review and users' attitudes toward the rated physician. We propose a model that relates review style and review number to constructs of review acceptance and check it with a Web-based experiment.

Methods: We employed a randomized 2x2 between-subject, factorial experiment manipulating the style of a physician review (factual vs emotional) and the number of reviews for a certain physician (low vs high) to test our hypotheses. A total of 168 participants were presented with a Web-based questionnaire containing a short description of a dentist search scenario and the manipulated reviews for a fictitious dental physician. To investigate the proposed hypotheses, we carried out moderated regression analyses and a moderated mediation analysis using the PROCESS macro 2.11 for SPSS version 22.

Results: Our analyses indicated that a higher number of reviews resulted in a more positive attitude toward the rated physician. The results of the regression model for attitude toward the physician suggest a positive main effect of the number of reviews (mean [low] 3.73, standard error [SE] 0.13, mean [high] 4.15, SE 0.13). We also observed an interaction effect with the style of the review—if the physician received only a few reviews, fact-oriented reviews (mean 4.09, SE 0.19) induced a more favorable attitude toward the physician compared to emotional reviews (mean 3.44, SE 0.19), but there was no such effect when the physician received many reviews. Furthermore, we found that review style also affected the perceived expertise of the reviewer. Fact-oriented reviews (mean 3.90, SE 0.13) lead to a higher perception of reviewer expertise compared to emotional reviews (mean 3.19, SE 0.13). However, this did not transfer to the attitude toward the physician. A similar effect of review style and number on the perceived credibility of the review was observed. While no differences between emotional and factual style were found if the physician received many reviews, a low number of reviews received lead to a significant difference in the perceived credibility, indicating that emotional reviews were rated less positively (mean 3.52, SE 0.18) compared to fact-oriented reviews (mean 4.15, SE 0.17). Our analyses also showed that perceived credibility of the review fully mediated the observed interaction effect on attitude toward the physician.

Conclusions: Physician-rating websites are an interesting new source of information about the quality of health care from the patient's perspective. This paper makes a unique contribution to an understudied area of research by providing some insights into how people evaluate online reviews of individual doctors. Information attributes, such as review style and review number, have an impact on the evaluation of the review and on the patient's attitude toward the rated doctor. Further research is necessary to improve our understanding of the influence of such rating sites on the patient's choice of a physician.

(J Med Internet Res 2015;17(4):e93) doi:10.2196/jmir.3991

KEYWORDS

physician reviews; physician-rating website; physician choice making; patient experiences; word of mouth

Introduction

Background

Online reviews are increasingly important sources of information for making different types of decisions. Recent industrial survey reports show that 90% of online shoppers read online reviews, and more than 80% of Internet users believe that these reviews affect their purchase behavior [1]. A growing number of people also share health care experiences online or rate the quality of their health care provider on physician-rating websites (PRWs) [2]. Thus, online resources, including advice from peers, are becoming a reliable source of health information, although the numbers still lag behind the numbers for commercial information searches [3]. The structure and content of PRWs are similar to other online rating websites in categories such as travel, hotels, or restaurants. PRWs typically provide information about a physician's address, phone number, office hours, and certifications. However, the most important feature of PRWs is their focus on the expression of patient opinion about, and satisfaction with, a physician's performance, which is online and visible to everyone. PRWs discuss the physician's standards by using user-generated data and reflect the physician's quality from the patient's point of view [4,5].

Although a growing number of Internet users are consulting online ratings or reviews of doctors or other health providers, we still know little about public physician-quality reporting and its impact on patient choice behavior [3,6]. Recent research in this field has investigated the number, distribution, and trend of evaluations on physician-rating websites [4,7]. In previously published studies, the percentages of rated physicians in the US varied from 16%—estimation for all physicians in the US in the period of 2005 to 2010 [3]—to 27%—analysis of 300 randomly selected Boston physicians on 33 US PRWs in 2009 [8]. A more recent analysis of physician ratings on the German PRW, jameda, found that 37% of all physicians in the German outpatient sector were rated on jameda in 2012 [9]. Findings regarding evaluation valence are quite consistent among different studies, showing that the vast majority of reviews are positive. For example, Lagu et al [8] reported that 88% of quantitative, and 89% of narrative, patient reviews were positive. Similarly, for the German PRW, jameda, it was shown that about 80% of all evaluations could be assigned to the two best rating categories (ie, *very good* or *good*) [9]. All in all, it seems that not only the number of PRW users have increased during the last years, but also the relevance of physician reviews for patients' decision processes.

Physician-Rating Websites and Patients' Evaluations of the Physicians

One topic which remains underresearched is the influence of PRWs when evaluating options and choosing a physician. More specifically, there are no studies that investigate the perception and effectiveness of online physician ratings and assess the impact of PRWs under experimental conditions [7]. We are aware of only one very recent study that examined the effect of the complexity of a choice set on consumers' choices of physicians in an experimental context and found that the quality of choice deteriorates as choice sets incorporate more options and more performance metrics [10]. Experiments allow researchers to estimate the causal effect of manipulable treatments to which experimental subjects are randomly allocated on a given outcome [11]—in the context of PRWs, for example, the effect of certain review characteristics on the patient's attitude toward the rated physician. There is also a research gap related to the content and nature of narrative reviews because most research in PRWs has focused on numerical ratings [12]. Recent investigations found that people not only evaluate ratings, but also scrutinize the written comments in online reviews [13]. Compared to simple numerical ratings (ie, star ratings or percentages), narrative reviews provide more detailed information and, thus, might add additional value to the decision-making process. Indeed, patients can write narrative commentaries in free text form on the vast majority of English-language and German-language PRWs [4,12,14]. However, the impact of narrative reviews depends on review characteristics such as review style and argument quality [15]. Consequently, this paper addresses this research gap and reports on an experimental study of the effects of review and context characteristics on patients' perceptions of the review, the reviewer, and the reviewed physician. In doing so, this paper contributes to a better understanding of the role of review and context characteristics in the evaluation of online physician reviews. More precisely, this study investigates how a more or less factually written online physician review, in conjunction with a low or high number of reviews for the rated physician, affects the attitude toward the rated physician, the perceived expertise and trustworthiness of the reviewer, and the perceived credibility of the review.

Determinants of Information Evaluation on Physician-Rating Websites

In recent years, considerable research has been directed at a better understanding of the effects of online rating sites on different aspects of choice behavior [16-20]. In these studies, different measures to investigate the effectiveness of online reviews have been used. In a comprehensive literature analysis of the impact of electronic word-of-mouth communication on consumer behavior [21], attitude, purchase intention, and product choice were identified as the most commonly investigated outcomes, followed by perceived usefulness, trust, and credibility constructs. Aspects of trust and credibility also have been shown to play a crucial role in the evaluation of online health information [22-24], and it can be assumed that they are at least equally relevant in the evaluation of physician ratings and reviews. Therefore, in this study we have chosen attitude toward the rated physician, and components of the message and source credibility as process and outcome variables.

Building on this line of work, information evaluation is considered a crucial determinant of the adoption of PRWs. According to traditional communication theories, a recipient's information evaluation results from an interaction of message characteristics—related to message content, comprising factors such as valence or information quality—source or context characteristics (eg, expertise, trustworthiness), and receiver characteristics (eg, previous experiences) [25]. From this range of factors that potentially determine the influence of peer reviews on people's decision making, in this study we have selected review style and number of reviews. As the effects of review valence have been extensively investigated by other researchers—with equivocal findings, though (eg, the review and synthesis of online word-of-mouth studies [26])—and as the majority of physician reviews on PRWs are positive [7,8,27], for this study we concentrate on the effects of positive reviews.

Review Style

A number of communication and persuasion researchers have investigated how various elements of messages, such as language intensity, style, and quality, influence message perceptions—see the overview by Eastin [23]. However, only a few studies have investigated the relevance of textual content and linguistic style in online reviews [15,17,26,28]. Using a dataset from the Amazon.com website and combining text mining with econometric techniques, researchers have demonstrated that the writing styles and language used in reviews determine both consumers' perceptions of those messages and their product choices [29]. It can be assumed that the style and the way reviews build up arguments for or against the rated physician also have an impact on how individuals evaluate the credibility of a review and the rated physician [30].

Consumer behavior literature offers contradictory findings about the influence of review style on review acceptance [18]. A qualitative study found that people prefer thematically structured, precise reviews that focus on simply describing the facts and refrain from extensively narrating the feelings of the reviewer [31]. That study defined perceived factuality as the conciseness, standardization, and specificity of an online review. In a similar vein, high-quality reviews are often operationalized as the ones that include relevant, comprehensive, and accurate product-related information [21,32]. In this respect, narrative and emotional expressions might be seen as a sign of subjectivity and, thus, reduce the perceived value of a review [33]. On the other hand, electronic peer-to-peer communication can be very influential because much of the information is presented in a narrative form and the emotional aspects of these narratives can be particularly persuasive [34]. A recent study found that reviews with emotional-laden elements, such as expressive slang and humor, were perceived as more valuable, at least up to a point [28]. Looking in the area of physician ratings, a recent context analysis found that narrative reviews for physicians sampled on the Yelp website, and in four English-speaking countries, lead to more usefulness ratings compared to pure fact-based reviews about the physician [35]. These heterogeneous arguments and prior findings prompt the consideration of some moderating variables when researching the influence of review style on review effectiveness. We assume that the influence of review style might interact with context factors, as further discussed, which might change if a more factually written review is more effective than a more emotionally written one.

Number of Reviews

We expect the volume of online reviews to be an indicator of the intensity of the underlying effect of peer-to-peer recommendations [36]. Previous theoretical and empirical research has found a positive relationship between the volume of word-of-mouth communication and product sales [37,38]. People tend to put more trust in the recommendations from large numbers of reviewers [39]. With a larger number of reviews, the reader might arrive at the conclusion that she/he can learn from, and more easily rely on, the positive experiences other people have made comments about [40]. In a very recent study on credibility assessment in online health forums, Lederman et al found crowd consensus to be a highly relevant criterion for the evaluation of experiential health information, as it provides a group opinion regarding the validity of an experiential statement [41]. Given the volume and dispersion of online information, recipients frequently use certain heuristics to evaluate messages [10,26]. In this vein, another basis for understanding the influence of review number and its interaction with review style on the evaluation of online reviews can be derived from dual-process theories [42,43]. From a dual-process perspective suggesting the co-occurrence of effortless and effortful processing modes in certain situations, the number of reviews for a physician can serve as a mental shortcut or heuristic cue that reduces the amount of time and cognitive effort needed to process the message [21,44]. A large number of reviews might act as a cue for crowd consensus that simplifies the decision process and leads to a less critical processing of the information content of a single review. In this way, a larger number of reviews for an individual doctor can enhance the value of narrative and emotional expressions. These reviews might otherwise be seen as a sign of subjectivity that can reduce the effectiveness of a review. Thus, we expect that with a growing number of reviews, the positive impact of factually written reviews on a recipient's information evaluation and the attitude toward the reviewed physician will become less salient.

Hypotheses

Hypothesis 1

The number of reviews for a physician moderates the impact of a factual review style on the attitude toward the reviewed physician. Specifically, the effect of a factual style on attitude toward the reviewed physician will be weaker if the number of reviews is high.

In addition to attitude toward the reviewed physician, we assume that the effects are processed via selected measures of source and message credibility as other relevant outcomes of review and context characteristics. Persuasion research outlines that evaluations of source and message credibility are dominant process variables that lead to attitude changes [45].

Receivers of eHealth information in general, and readers of online physician reviews in particular, ought to be able to assess the credibility of the source and the message as important steps to information processing [46]. Credibility is the believability of a source or message [47] and has emerged as a critical indicator of eHealth information quality [48]. The components of source credibility that have been commonly identified are trustworthiness and expertise [49]. Trustworthiness refers to an information provider's intention to tell the truth or give unbiased information, and expertise refers to the extent to which an information provider is perceived capable of making correct assertions. Message credibility typically examines how message or information characteristics influence perceptions of believability. Major factors influencing message credibility include message structure, content and language, and plausibility of arguments [25,47].

Usually, writers of reviews stay anonymous and it is difficult for readers to assess source credibility and to determine if the information they receive is trustworthy [50]. Credibility research on media suggests that limited knowledge of source competence causes respondents to seek message-inherent heuristic cues (eg, language style and review number) to evaluate the information [23,51]. Cues embedded in the review and its presentation become characteristics with which to evaluate the message validity. In the context of this study, we examine the impact of review style and review number on both perceived credibility of the source (ie, expertise and trustworthiness of the reviewer) and the message. It is expected that more factually written physician reviews will induce higher levels of perceived expertise and trustworthiness of the reviewer and will lead to higher review credibility, and these effects will be moderated by the number of reviews.

Hypothesis 2

The number of reviews for a physician moderates the impact of a factual review style on the perceived expertise of the reviewer. Specifically, the effect of a factual style on perceived expertise of the reviewer will be weaker if the number of reviews is high.

Hypothesis 3

The number of reviews for a physician moderates the impact of a factual review style on the perceived trustworthiness of the reviewer. Specifically, the effect of a factual style on perceived trustworthiness of the reviewer will be weaker if the number of reviews is high.

Hypothesis 4

The number of reviews for a physician moderates the impact of a factual review style on the perceived credibility of the review. Specifically, the effect of a factual style on perceived credibility of the review will be weaker if the number of reviews is high.

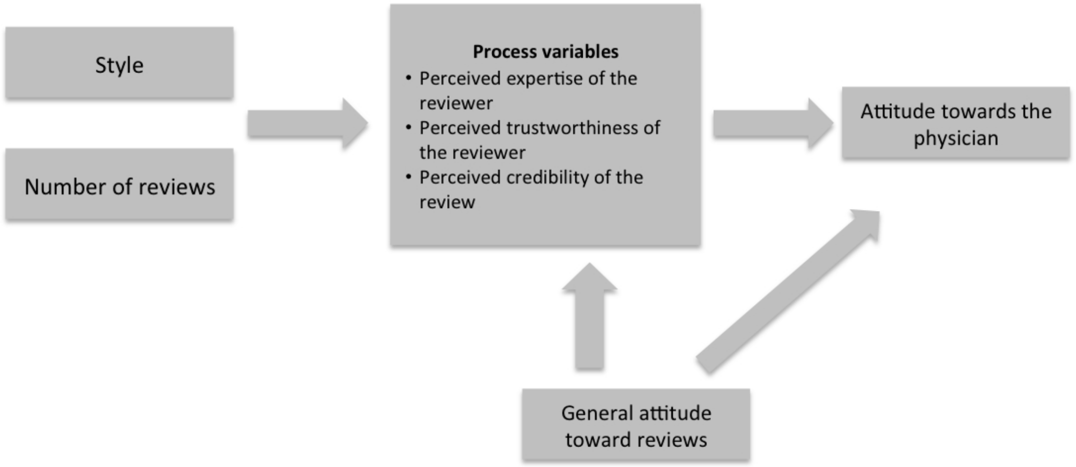
Already, early studies in credibility research on media [52] found that the “trustworthiness” of a source significantly affected both the acceptance of the presented material and changes in opinion and attitude [23]. In the consumer behavior literature, dimensions of trust have been described as key mediating variables leading to positive attitudinal or behavioral relationship outcomes [53,54]. Following these lines of research, we also investigated the mediating effects of the credibility variables on the attitudinal outcome variable and suggest the following hypothesis.

Hypothesis 5

Perceived expertise and trustworthiness of the reviewer, as well as perceived credibility of the review, will mediate the interaction effect of style and review number on ratings of attitude toward the rated physician.

Figure 1 summarizes the proposed model of how the style of the review and the number of reviews received influence attitude formation toward a physician.

Figure 1. Influence of style and the number of reviews on attitude toward the reviewed physician.



Methods

Overview

To investigate the proposed hypotheses, we conducted a 2x2 online experiment—two review styles (emotional vs factual) by two numbers of reviews (low vs high)—between subjects, with 168 participants.

Design

With careful isolation of the variables under consideration, the aim was to obtain an experimental design that allows for estimating the effects of review style and number. A search scenario for a dentist was chosen as a medium-to-high-involvement health service setting that seemed appropriate for a doctor choice that, in reality, could be primarily based on online reviews and would not allow for seeking personal word-of-mouth recommendations. Therefore, the scenario for an urgent need to consult a dentist because of a toothache shortly after relocation to a new city was created (see the description below). The choice of a general practitioner, for example, in the case of a flu, presumably could be based rather on practical decision criteria such as a short distance from the place of residence to the doctor's office. On the other hand, in a situation where surgery is required it can be assumed that patients would not exclusively rely upon information provided on anonymous PRWs, but look for additional information from someone they personally know. Moreover, as both male and female participants were addressed, gynecologists and obstetricians were ruled out. On a final note, a search scenario for a dentist was considered more relevant for more people compared to, for instance, a search scenario for other specialists, such as orthopedists, internists, or pediatricians.

The doctor search scenario in the questionnaire was described as follows (in Austria, the term physician is used for dentists as well):

> *Please place yourself in the situation and imagine what this would be like for you:*
>
> *You recently have moved to another city. As you suddenly have a racking toothache you start searching for a new dentist. Unfortunately, you do not know yet any dentist in this city and also cannot draw on recommendations from friends or acquaintances. Therefore, you decide to look for a dentist in your surroundings on the Web and read through online physician evaluations on this occasion.*
>
> *You enter your search criteria on the Austrian physician rating website www.docfinder.at. On the next webpage, you are shown the profile of a dentist and the corresponding physician evaluation. Please read both descriptions carefully and then respond to the questions below.*

The style of the reviews was manipulated by showing two versions of an online review for a fictitious dental practitioner on docfinder, the most popular PRW in Austria. The two review versions for the fictitious dentist, Dr Frank Weber, were created after a thorough and comprehensive analysis of the content, scope, and style of different reviews for dentists and other physicians on docfinder and other PRWs in Austria and Germany to keep them as realistic as possible. The reviews showcased a fictitious dentist visit and addressed the same order of topics in both versions. Moreover, care was taken to make sure the topics addressed in the two review versions could be assigned to some of the most relevant categories of patient concerns that have been identified in previous literature on narrative comments on PRWs, for example, as in Emmert et al [12]. One review was written in a more emotional way (eg, with the heading "fantastic doctor" and the use of emoticons such as a smiley, capital letters, and exclamation marks to strengthen the intensity of the positive message), while the other headline and the content were centered around facts about the physician's service and the practice (eg, with the heading "modern technology and competent advice" and with only positive and concrete comments, no emoticons, no capital letters, and no exclamation marks). Both the fact-oriented and the emotional reviews included comments related to the professional competence, the efforts of the physician and the office staff, the practice equipment, and the waiting time to get an appointment. Friendliness of the physician and staff was not explicitly included in the fact-oriented review because it was considered to be more of an emotional attribute.

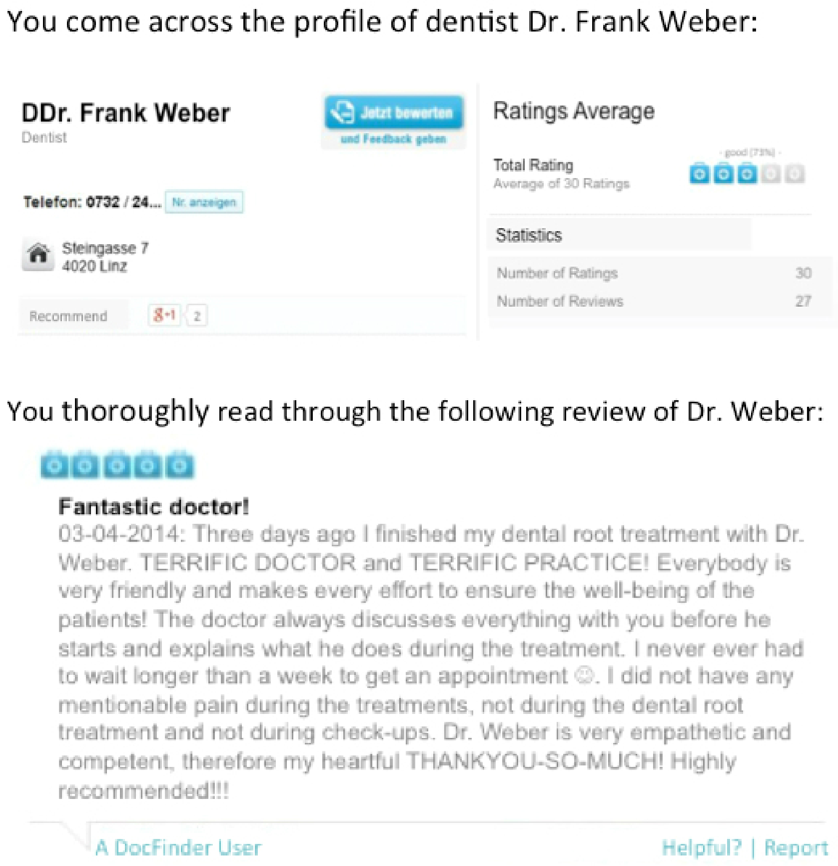
To check whether our manipulations worked, we asked how emotional or factual the review was perceived on a 7-point semantic differential, ranging from 1 (*emotional*) to 7 (*factual*). A *t* test for mean comparison showed a significant difference (t_{164}=7.470, *P*<.001). Emotional reviews were rated as 2.79 (SD 1.32) while the factual reviews were rated as 4.47 (SD 1.55).

The number-of-reviews-received factor was manipulated by presenting two different versions of the physician profile for Dr. Frank Weber. Both versions of the profile included the physician's address, telephone number, and the average overall rating of the physician—3 doctor's cases out of 5 (*good* [73%])—in both versions. On docfinder, the average overall physician rating is shown with doctor's case symbols (ie, icons that look like doctor's medical bags). Ratings range from 0 doctor's cases, meaning *insufficient*, to 5 doctor's cases, meaning *excellent*. The manipulated review itself was rated as positive with 5 doctor's cases out of 5. All this basic information was equal in both profile versions. Only the provided statistical information (ie, number of ratings and number of reviews received) was different in the two number-of-reviews-received conditions. The low number was 3 for both ratings and reviews received, while the high number was 30 for ratings and 27 for reviews received. Again, these numbers resulted from searching different physicians and noting how many reviews they received. On docfinder, in terms of review numbers, dentists considered the "top physicians" received about 10 times more reviews compared to the bottom-10%-rated physicians, who received the lowest number of reviews.

A 7-point scale assessing the perceived number of reviews, ranging from 1 (*few*) to 7 (*many*), served as a manipulation check. It showed that, generally, the 3 ratings and 3 reviews (low number) as well as the 30 ratings and 27 reviews (high number) were perceived as *few* to *average* number of reviews, but with a clear and significant difference between them (t_{164}=8.889, *P*<.001). In the low number conditions, the mean perceived number of reviews was rated as 1.98 (SD 1.12), while the high number was rated as 3.87 (SD 1.64).

In the following figures, the four experimental conditions are illustrated. The experimental report card of the physician profile above is for the manipulation of the number of received ratings (low vs high), while the review text below is for either the fact-oriented or emotional style manipulation. English translations of two report cards are provided in Figure 2 and Figure 3—the originals were in German.

Figure 2. Manipulated rating card for the high-number/emotional review condition.



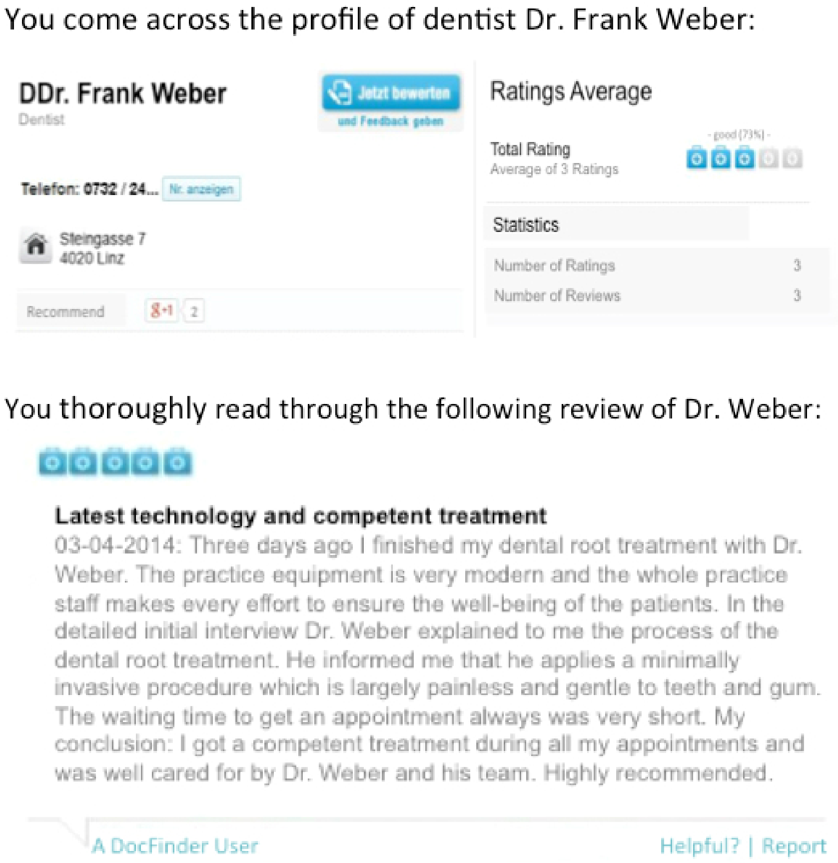
You come across the profile of dentist Dr. Frank Weber:

You thoroughly read through the following review of Dr. Weber:

Fantastic doctor!

03-04-2014: Three days ago I finished my dental root treatment with Dr. Weber. TERRIFIC DOCTOR and TERRIFIC PRACTICE! Everybody is very friendly and makes every effort to ensure the well-being of the patients! The doctor always discusses everything with you before he starts and explains what he does during the treatment. I never ever had to wait longer than a week to get an appointment ☺. I did not have any mentionable pain during the treatments, not during the dental root treatment and not during check-ups. Dr. Weber is very empathetic and competent, therefore my heartful THANKYOU-SO-MUCH! Highly recommended!!!

Figure 3. Manipulated rating card for the low-number/factual review condition.



You come across the profile of dentist Dr. Frank Weber:

You thoroughly read through the following review of Dr. Weber:

Latest technology and competent treatment

03-04-2014: Three days ago I finished my dental root treatment with Dr. Weber. The practice equipment is very modern and the whole practice staff makes every effort to ensure the well-being of the patients. In the detailed initial interview Dr. Weber explained to me the process of the dental root treatment. He informed me that he applies a minimally invasive procedure which is largely painless and gentle to teeth and gum. The waiting time to get an appointment always was very short. My conclusion: I got a competent treatment during all my appointments and was well cared for by Dr. Weber and his team. Highly recommended.

Measures

Established scales were used and adapted to a patient-doctor context, where needed, to measure each of the investigated constructs. The questionnaire was pretested on two experts and 10 other people in order to identify possible problems in terms of clarity and accuracy, and also to check to see if the created versions of the review and profile for the dentist were perceived as realistic. Thereafter, a few changes were made in order to improve the presentation order of the items, based on comments and feedback. Three items adapted from Hwang et al [55] gauged the attitude toward the physician using a 7-point scale (mean 3.94, SD 1.31, alpha=.942). Five items each, taken from Ohanian et al [56] assessed perceived expertise of the reviewer (mean 3.55, SD 1.26, alpha=.931) and perceived trustworthiness of the reviewer (mean 4.09, SD 1.29, alpha=.959), all of which were measured on 7-point agreement scales. Two items—the review is trustworthy and the review portrays a realistic picture of the physician—using a 7-point agreement scale [57], measured the perceived credibility of the review (mean 3.85, SD 1.25, alpha=.853). It can be assumed that any of the review-related perceptions of online reviews are influenced by how often people use reviews in general, and what attitude they have toward using them. Thus, we measured the general attitude toward online reviews (mean 4.37, SD 1.56, alpha=.934) with three items [58] and tested whether the slope direction was different between the groups. This check yielded no significant difference and, therefore, general attitude toward reviews was included as a covariate in the analyses to control for this influence. A PDF version of the questionnaire—German original, plus English translation—is provided in Multimedia Appendix 1.

Sample

A total of 168 individuals clicked on the link to the online experiment and filled in the survey. Out of all the surveys, 2 cases were incomplete and could not be integrated in the analyses. Out of 166 participants, 50 (30.1%) were male. The average age was 27.7 years (SD 7.8), with a range from 16 to 58 years. Of 166 participants, 61 (36.7%) were full-time students, 60 (36.1%) were professionals, 35 (21.1%) reported that they were both studying and working, and 10 (6.0%) reported some other employment status. We checked whether age, gender, or occupation showed any significant differences among the randomized groups. Analysis of variance (ANOVA) and chi-square tests yielded no significant differences. Table 1 summarizes the distribution of occupation and gender among the randomized groups, as well as the average age of participants per experimental condition.

Only 7.2% (12/166) indicated that they never consult online reviews in general. However, many of the participants (75/166, 45.2%) have never written an online review, and 34.9% (58/166) indicate that they have written between one and three online reviews. Generally, the adoption of online reviews for health services compared to those for products and services is quite low, as 53.0% (88/166) of the participants have never consulted a physician-rating website before. Only 6.0% (10/166) have previously written a physician review.

Table 1. Characteristics of the sample (n=166).

Characteristics	Total, n (%) or mean (SD)	Experimental group, n (%) or mean (SD)			
		Low-number/ factual review (n=42)	Low-number/ emotional review (n=41)	High-number/ factual review (n=44)	High-number/ emotional review (n=39)
Gender, n (%)					
Male	50 (30.1)	18 (43)	10 (24)	13 (30)	9 (23)
Female	116 (69.9)	24 (57)	31 (76)	31 (70)	30 (77)
Occupation, n (%)					
Student	61 (36.7)	17 (40)	17 (41)	17 (39)	10 (26)
Professional	60 (36.1)	18 (43)	10 (24)	17 (39)	15 (38)
Working and studying	35 (21.1)	5 (12)	12 (29)	8 (18)	10 (26)
Other	10 (6.0)	2 (5)	2 (5)	2 (5)	4 (10)
Age in years, mean (SD)	27.7 (7.8)	27.1 (6.3)	28.4 (8.4)	28.1 (8.4)	27.0 (8.0)

Procedure

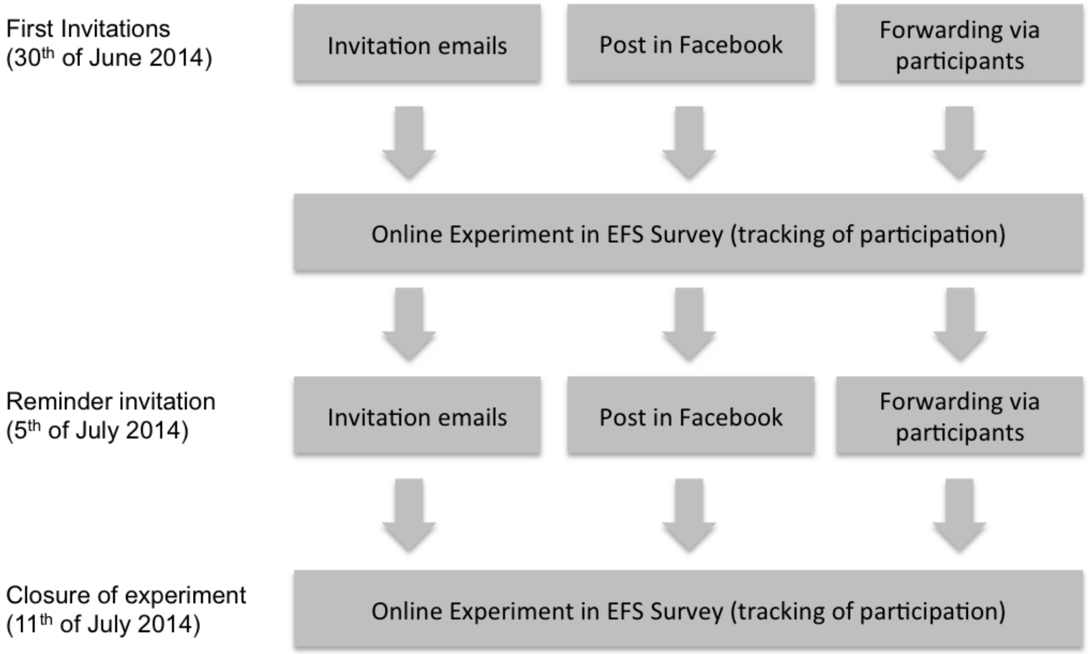
The experiment was conducted online, using EFS Survey from Questback. The participants were recruited via email and social media. Undergraduate and graduate students at a mid-sized European university were invited to participate via an email message sent to their university email accounts. In addition, the link to the online survey was posted on the Facebook pages of the Department of Marketing and International Management and those of some of the research team members. Participants were also encouraged to actively forward the link to their friends.

The questionnaire was available online from June 30 to July 11, 2014. In total, 264 participants opened the link to the experiment, 168 (63.6%) participants finished the survey, and 166 (62.9%) of all the surveys contained complete data on all measures (Figure 4).

First, respondents were asked some basic demographic questions. They were also asked about their use of online reviews in general, and online physician reviews in particular. They were then shown one of the four review conditions, which were randomly assigned by the software. Out of the 166 participants, 41 (24.7%) were exposed to the low-number-of-reviews/emotional condition, 42 (25.3%) to the low-number-of-reviews/factual condition, 39 (23.5%) to the high-number-of-reviews/emotional condition, and 44 (26.5%) to the high-number-of-reviews/factual condition. Right after participants read the review reports, we assessed the perceived credibility, followed by manipulation checks for review style, review number, and the attitude toward the dentist. Finally, attitudes toward review reports in general were recorded. After finishing the survey, if participants wanted, they could have their names entered into a raffle to win one of three €20 gift certificates.

To investigate the proposed hypotheses 1 to 4, we computed moderated regression analyses using the PROCESS macro 2.11 from Hayes for SPSS version 22 [59]. For testing the effects on the dependent variables, we applied the following settings: Model 1 (5000 bootstraps, confidence level of 95%), with the review style as the independent variable and the number of reviews as moderator, was estimated. The attitude toward online reviews in general was entered as a covariate to control for its influence in all of the analyses. Bootstrapping provides upper and lower level confidence intervals (ULCI and LLCI, respectively). If the range of these two does not include zero, the analysis shows significance. To investigate hypothesis 5, we calculated a moderated mediation model in PROCESS—Model 8: 2x2 design with two main effect variables and one interaction term—with perceived credibility of the review, perceived trustworthiness of the reviewer, and perceived expertise of the reviewer as mediators, and the two manipulated factors style and number as independent variables. Attitude toward the physician served as the outcome variable, with 5000 bootstraps and the three process variables—perceived expertise and trustworthiness of the reviewer, and perceived credibility of the review—as mediators. The number of reviews received, as well as the review style, were entered as independent variables, and general attitude toward reviews was entered as a covariate.

Figure 4. Recruitment process.



Results

Hypothesis 1 proposed a positive effect of factual review style, which should be more prevalent if the physician received only a few reviews. The results of our regression model for attitude toward the physician suggest a positive main effect of the number of reviews (unstandardized regression coefficient [R]=0.398, t_{161}=2.111, 95% CI 0.026-0.768, *P*=.04). Thus, when the dentist had more reviews, the attitude toward him was more positive (mean 4.15, standard error [SE] 0.13), compared to the condition where the dentist only had a few reviews (mean 3.73, SE 0.13). Besides this interesting main effect, the interaction effect of review style and number of reviews received was also significant (R=-0.798, t_{161}=2.118, 95% CI -1.541 to -0.054, *P*=.04). As the effect was negative, this implies that the positive effect of a factual review style is mitigated when a high number of reviews is present.

We conducted single comparisons with conditional effect analyses, which revealed that the emotional review style lead to a less positive attitude (mean 3.44, SE 0.19) compared to the factual review style (mean 4.09, SE 0.19), if only a small number of reviews were received (95% CI 0.132-1.179, *P*=.01). If the physician received many reviews, factual and emotional review style did not result in different attitude ratings of the dentist. Furthermore, an emotional review style in the low-number-of-reviews-received condition also lead to a significantly less favorable attitude rating (mean 3.44, SE 0.19) compared to emotional reviews in the high-number-of-reviews-received condition (mean 4.23, SE 0.19) (95% CI 0.263-1.329, *P*=.004). The impact of the factual review style was not different in the two number-of-reviews-received groups. Figure 5 shows the interaction effect. In summary, factually written reviews performed better compared to emotional ones if the number of reviews received was low. However, if the physician had already received many reviews, the emotionally written review resulted in a more positive attitude. The covariate, general attitude toward online reviews, positively influenced the attitude toward the dentist (R=0.288, t_{161}=4.861, 95% CI 0.179-0.423, *P*<.001). Thus, hypothesis 1 was confirmed.

In hypotheses 2 to 4, we predicted a positive effect of a factual review style on dimensions of source and message credibility. We also predicted that this effect would be moderated by the number of reviews received, as we expected that if only a few ratings and reviews were received, fact-oriented information would be more important. For each of the three variables, we calculated a separate moderated regression as we did for the test of hypothesis 1. The summaries of all these single regression models are displayed in Tables 2-4.

When testing hypothesis 2, we found a main effect of review style on perceived expertise of the reviewer. More factually written reviews (mean 3.90, SE 0.13) lead to a stronger perception that the reviewer had some expertise compared to the emotionally (mean 3.19, SE 0.13) written ones (R=0.710, t_{161}=3.882, 95% CI 0.349-1.071, *P*<.001). All other effects were not significant, thus, hypothesis 2 is only supported for the main effect of review style (see Table 2).

When testing hypothesis 3 for the influence of review style and the number of reviews received on perceived trustworthiness of the reviewer, the moderated regression analyses yielded no significant effects, except for the covariate attitude toward reviews (R=0.207, t_{161}=3.237, 95% CI 0.081-0.333, *P*<.001) (see Table 3). Thus, hypothesis 3 had to be rejected. However, the direction of the effect was similar to that for perceived credibility of the review. For the high-number-of-reviews-received condition, emotional and factual reviews lead to similar levels of perceived trustworthiness of the reviewer (mean 4.23, SE 0.20, mean 4.12, SE 0.19, respectively). However, for the low-number-of-reviews condition, the difference in the measure for perceived trustworthiness of the reviewer tended to be larger (mean 3.76, SE 0.20, mean 4.26, SE 0.20, respectively). Although not significant, the direction of the interaction effect did not contradict our other findings.

When testing hypothesis 4 for the influence of review style and the number of reviews received on perceived credibility of the review, the moderated regression analysis confirmed the proposed interaction effect (R=-0.852, t_{161}=-2.419, 95% CI -1.547 to -0.156, *P*=.02) (see Table 4). Single comparisons showed that in the condition where the physician received only a few reviews, factual reviews (mean 4.15, SE 0.17) were perceived as more trustworthy than emotional reviews (mean 3.52, SE 0.18, 95% CI 0.139-1.117, *P*=.01). Also, the comparison between emotional reviews in the low- and high-number-of-reviews-received conditions was significant. In the condition of many reviews received, emotionally written reviews resulted in higher credibility judgments (mean 4.04, SE 0.18) compared to emotional reviews in the condition of few reviews received (mean 3.52, SE 0.18, 95% CI 0.013-1.009, *P*=.045). Thus, hypothesis 4 is confirmed by our data. Figure 6 displays the interaction effect.

Hypothesis 5 proposes that the effects of review style and review number are mediated via the process variables perceived trustworthiness and expertise of the reviewer, as well as the perceived credibility of the review. Since we found significant main and interaction effects in the single analyses, we carried out a moderated mediation analysis as proposed by Hayes [59]. According to this analysis, it is essential that the effects from the independent variables on the mediators are established first. Furthermore, a relationship between the mediator and the dependent variable needs to be established—that is, a mediated indirect effect. Yet, nonmediated direct effects of the independent variables on the dependent variable are also possible. Thus, a mediation analysis must check to see if direct and indirect effects are observable [60]. To check this, the modeling tool PROCESS calculates a set of regression analyses. In the first two steps, the mediators and the dependent variables are regressed on the independent variables. Then, in a total model all independent variables (and possible covariates), as well as the mediator variables, are entered as predictors for the ultimate dependent variable. At the end, bootstrapping determines if the direct and indirect effects are significant. If both yield significance, we have partial mediation. If only indirect effects are significant, full mediation is observed, and if only direct effects are observed, no mediation occurs. To test the moderated mediation, bootstrapping is also applied. If the indirect effect of the highest order interaction bootstrapping results of LLCI and ULCI does not include zero in its range, a significant mediation effect is present.

Tables 2-5 summarize the results for the mediated regression analyses. The results for the regressions for the three mediators—hypotheses 2 to 4—showed a positive main effect of style on the perceived expertise of the reviewer, indicating that more factual reviews lead to higher expertise ratings. Furthermore, for credibility of the review, we found a significant interaction effect of review style and number of reviews received. The final total model includes all independent and mediation variables. The total effects of the independent variables, both indirect via the mediators and direct on the dependent variable, were calculated and displayed in Table 5. We followed the steps proposed by Zhao et al [59] to determine full or partial mediation and calculated the total regression model.

The total regression model showed a good predictive power (R^2=0.62). When the mediators were entered into the model, attitude toward the physician was significantly influenced by the perceived trustworthiness of the reviewer (R=0.359, t_{161}=5.101, 95% CI 0.220-0.498, *P*<.001), as well as by the perceived credibility of the review (R=0.430, t_{161}=5.592, 95% CI 0.278-0.582, *P*<.001). Furthermore, the main effect of review number was also significant in the total model (R=0.281,

t_{161}=2.153, 95% CI 0.023-0.538, *P*=.03). The analyses confirmed that the influence of review style and number of reviews received was fully mediated via credibility of the review. This is seen as the direct effect becomes insignificant, and an indirect effect of the interaction term on the dependent variable, solely mediated via credibility of the review, is found (R of the indirect effect=-0.367, 95% CI -0.788 to -0.082, *P*=.05). Although we found a significant effect for the review style on the perceived expertise of the reviewer, the regression analyses showed that perceived expertise of the reviewer was not related to the attitude toward the physician, therefore, the main effect was not mediated. Thus, the proposed mediation effect in hypothesis 5 was only established for perceived credibility of the review, which fully mediated the interaction effect. General attitude toward reviews had a positive significant effect on all mediators, as well as on the dependent variable attitude toward the physician, indicating that people who are more positive toward reviews in general, show higher ratings in all outcomes.

In summary, the mediation analysis showed that perceived trustworthiness of the reviewer and credibility of the review influenced the attitude toward the doctor, as did the number of reviews received. However, the review style also had an influence. Our analysis suggests that factual reviews lead to a more positive credibility evaluation of the review, but only in the low-number-of-reviews-received condition. Due to the positive effect of review credibility on attitude toward the physician, the effect of style also transfers to the attitude toward the physician. The data partially support the hypothesis of a positive main effect of a fact-oriented review style. A more fact-oriented review style has a positive effect on the perceived expertise of the reviewer. However, perceived expertise of the reviewer was not a significant predictor for attitude toward the reviewed doctor. See Tables 2-5 for summaries of the above analyses.

Table 2. Results of the moderated mediation analyses for the outcome variable, perceived expertise of the reviewer.

Source[a]	R[b]	SE[c]	*t*	*P*	95% CI
Constant	2.798	0.278	10.070	<.001	2.249 to 3.347
Review style	0.710	0.183	3.882	<.001	0.349 to 1.071
Review number	0.289	0.182	1.586	.115	-0.071 to 0.648
Review style x review number	-0.293	0.365	-0.804	.422	-1.013 to 0.427
Attitude toward online reviews	0.169	0.060	2.826	.005	0.051 to 0.287

[a] Model summary: R=0.396, R^2=0.157, $F_{4,161}$=7.486, *P*<.001.

[b] Unstandardized regression coefficient (R).

[c] Standard error (SE).

Table 3. Results of the moderated mediation analyses for the outcome variable, perceived trustworthiness of the reviewer.

Source[a]	R[b]	SE[c]	*t*	*P*	95% CI
Constant	3.181	0.297	10.705	<.001	2.595 to 3.768
Review style	0.199	0.196	1.015	.312	-0.188 to 0.585
Review number	0.158	0.195	0.813	.417	-0.226 to 0.542
Review style x review number	-0.615	0.390	-1.577	.117	-1.385 to 0.155
Attitude toward online reviews	0.207	0.064	3.237	.002	0.081 to 0.333

[a] Model summary: R=0.294, R^2=0.086, $F_{4,161}$=3.800, *P*=.006.

[b] Unstandardized regression coefficient (R).

[c] Standard error (SE).

Table 4. Results of the moderated mediation analyses for the outcome variable, perceived credibility of the review.

Source[a]	R[b]	SE[c]	*t*	*P*	95% CI
Constant	2.520	0.268	9.390	<.001	1.990 to 3.050
Review style	0.202	0.177	1.144	.255	-0.147 to 0.551
Review number	0.085	0.176	0.484	.629	-0.262 to 0.432
Review style x review number	-0.852	0.352	-2.419	.017	-1.547 to -0.156
Attitude toward online reviews	0.309	0.058	5.353	<.001	0.195 to 0.422

[a] Model summary: R=0.429, R^2=0.184, $F_{4,161}$=9.085, *P*<.001.

[b] Unstandardized regression coefficient (R).

[c] Standard error (SE).

Table 5. Results of the moderated mediation analyses for the outcome variable, attitude toward the physician.

Source[a]	R[b]	SE[c]	*t*	*P*	95% CI
Constant	0.246	0.283	0.869	.386	-0.314 to 0.806
Expertise of the reviewer	0.079	0.059	1.329	.186	-0.038 to 0.196
Trustworthiness of the reviewer	0.359	0.070	5.101	<.001	0.220 to 0.498
Credibility of the review	0.430	0.077	5.592	<.001	0.278 to 0.582
Review style	0.043	0.136	0.314	.754	-0.226 to 0.311
Review number	0.281	0.130	2.153	.033	0.023 to 0.538
Review style x review number	-0.187	0.264	-0.710	.479	-0.708 to 0.334
Attitude toward online reviews	0.068	0.046	1.461	.146	-0.024 to 0.159

[a] Model summary: R=0.786, R^2=0.617, $F_{7,158}$=36.392, *P*<.001.

[b] Unstandardized regression coefficient (R).

[c] Standard error (SE).

Figure 5. Attitude toward the physician as a function of review style and number of reviews.

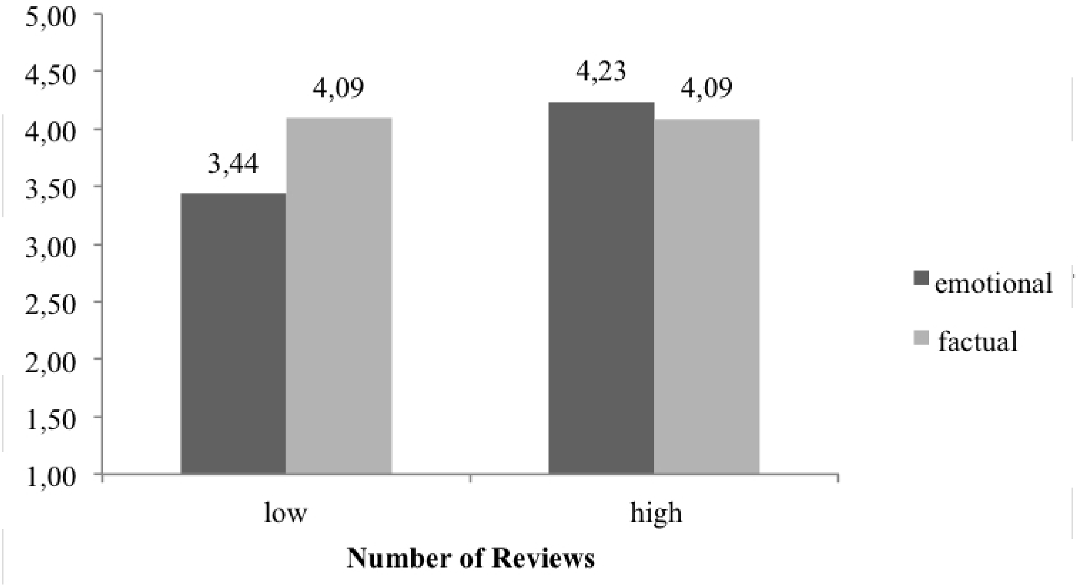
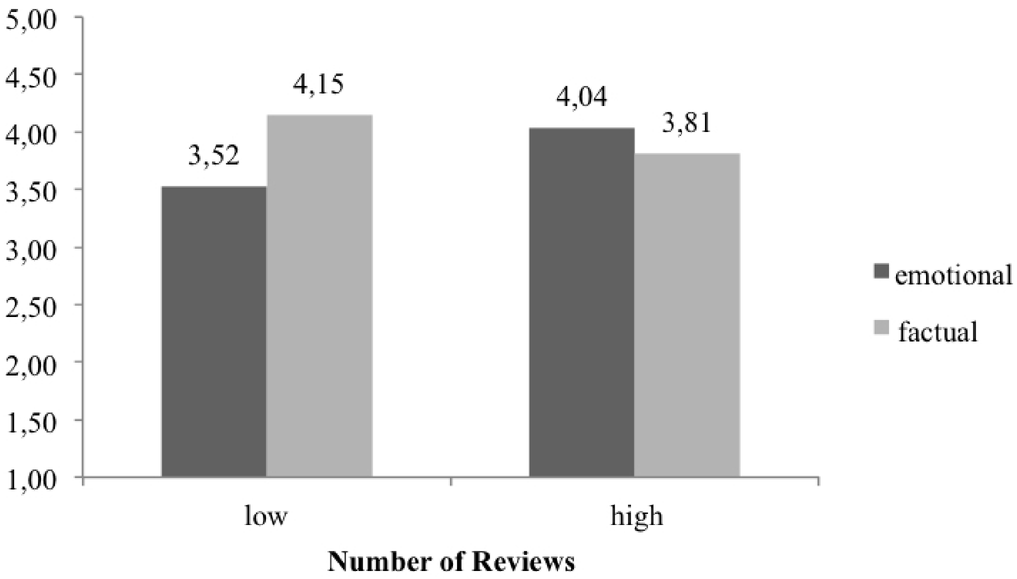


Figure 6. Credibility of the review as a function of review style and number of reviews.



Discussion

Principal Findings

The Internet has become an indispensable information medium for health services during the last years. Just as on other newly ubiquitous rating sites, an increasing number of people comment on their health care experiences online and rate the quality of their health care provider on physician-rating sites [2]. PRWs provide insight into the quality of care from the patient's perspective and are a quick and easily accessible information source for patients who are seeking a physician. However, little is known about how physician reviews and ratings might affect physician choice making. As Schlesinger et al pointed out in a recent paper, consumer choice of doctors remains a "black box" that research has scarcely illuminated [10]. In their exploratory experimental study they investigate the impact of the complexity of information provided on PRWs on the quality of patients' choice of doctors. Beyond that, with the growing use of PRWs, it is essential to identify which factors are most relevant for decision-making processes in this context [10]. One step in this direction is to improve our understanding of how patients develop attitudes toward doctors based on online reviews. Drawing on research from services as experience goods, we proposed a mediating impact of the expertise and trustworthiness of the reviewer, as well as the credibility of the review, on the attitude formation toward the physician. Furthermore, we expected that these relationships would depend on how the review is written and how many reviews the physician received.

We found that a more factual review style was positively related to the perceived expertise of the reviewer. This finding is in agreement with the findings of Lee and Koo [18] who showed that the credibility of online reviews with objective information was higher than that of online reviews with more emotional, subjective statements. The present findings also seem to be consistent with other research in the context of consumer products, which found that a more factual style is beneficial for the evaluation of the reviewer's expertise and important for the generation of trust in the review [61]. Yet, in the current study the positive effect on perceived expertise as a dimension of source credibility did not translate into a more positive attitude toward the physician. And while trustworthiness of the reviewer positively influenced the attitude toward the physician, this effect was not dependent on the proposed independent factors investigated in our experiment. However, this study found that factual reviews lead to a more favorable attitude toward the physician only if a low number of reviews were received. We also found that the emotional reviews had greater impact when the physician received a high number of reviews. This interaction effect on attitude toward the physician was fully mediated by credibility of the review. Furthermore, we found a direct effect of the number of reviews received, which suggests that physicians were evaluated more favorably when they received more reviews. These findings corroborate previous research that found that physicians who received a higher number of ratings were shown to have better ratings [9].

Implications

This leads to some important implications from the practitioner's point of view. Firstly, it could be beneficial for physicians to obtain more reviews on their services. Similar suggestions from services marketing how to encourage patients to write reviews could be applied, like invitations or signage in the waiting area. Another approach to engage patients in writing reviews might work through the PRW itself. Similar to hotel booking services, if, for example, the contact to the physician is established via a PRW contacting option, the PRW could follow up with a reminder email to review the visit to the doctor.

Secondly, if only a few reviews are present and those are mainly emotional, even if they're not negative, this leads to a less favorable attitude toward the physician and possibly reduces the probability of choosing this doctor. Yet, a physician could add some value here by responding to the reviews in a fact-oriented manner. This can easily be done if a feedback loop on PRWs is provided that allows physicians to respond to patients' comments [12].

On the other hand, if a doctor received many reviews, a mix of emotional and factual reviews could be even more beneficial. This could be achieved by engaging different types of patients to write reviews. Also, these different types of reviews might attract further users of the PWR as the reviews might be perceived as more helpful if they contain different style elements. A recent content analysis on reviews for doctors on Yelp found that narrative reviews seemed to generate more usefulness ratings compared to short, pure fact-based ones [35].

Another minor finding from this study is the relatively low usage of health-related reviews compared to other product and service categories. While about 46% of our respondents reported that they have already consulted online reviews when searching for a physician, only 6% actually have written and posted a physician review themselves. These numbers are different from those reported by Emmert et al [7], who consulted a German online panel in January 2013 and found that only about 25% of all respondents had used a website when searching for a physician, but 11% had already posted a rating on a PRW. The differences could either be related to regional differences in the diffusion of PRWs between Germany and Austria or be a sampling effect because our experimental study relied on a convenience sample, as most experimental research does.

Taken together, the results of this research support the idea that more reviews are beneficial for a physician, as a high number of reviews might indicate her/his reputation as a sought-after practitioner. Our research also shows that people rate physicians more favorably as a result of both factual and emotional reviews if they have a wide array of reviews to choose from. However, in a situation of scarce information, more fact-oriented reviews lead to more favorable attitudes toward the reviewed physician.

Limitations

There are some limitations to our study. First, we exposed the participants to only one physician review to see the effect of a more emotional or more fact-oriented review style, yet that did not necessarily reflect a real-world setting where comparisons and long lists of reviews are present. Moreover, in order to focus on the effect of review style and review number on attitude toward the rated physician and review and reviewer credibility, other characteristics of physician reviews, such as valence, length, review themes, and overall ratings, were controlled for. Future research could explore whether changes in these dimensions alter the effect of review style and number on review acceptance. Besides, as an online experiment was employed we could not control for distraction during the experiment. Another limitation is the exclusive focus on dentists. An investigation of the relevance of online reviews for the choice of other specialists or general physicians would not have to consider exactly the same topics or patients' concerns, which also might have an effect on the influence of these reviews—future studies should address this point. Another limitation of our study is that we included only positive reviews. Therefore, testing the effect of review valence might be a valuable extension of the work. Finally, important limitations concern the nonprobability sampling technique and the narrowly focused online sample, which limited our ability to generalize the study's findings to a broader population of patients. As the average age of our sample was only 27.7 years, the findings in particular provide some insights into how this younger age group evaluates online reviews of individual doctors. Further research might address the impact of age differences on the usage behavior of PRWs.

Acknowledgments

The authors would like to thank Daniela Burgstaller and Sabrina Oberressl who created the review versions in their master's theses and collected the data.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire used to measure investigated constructs—German original and English translation.

[PDF File (Adobe PDF File), 996KB - jmir_v17i4e93_app1.pdf]

References

1. Zhang KZ, Cheung CM, Lee MK. Examining the moderating effect of inconsistent reviews and its gender differences on consumers' online shopping decision. Int J Inf Manage 2014 Apr;34(2):89-98. [doi: 10.1016/j.ijinfomgt.2013.12.001]
2. Verhoef LM, Van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014 Feb;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
3. 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]
4. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013 Feb;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]

5. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014 Mar;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
6. 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]
7. 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 Aug;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
8. 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]
9. 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 Aug;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
10. Schlesinger M, Kanouse DE, Martino SC, Shaller D, Rybowski L. Complexity, public reporting, and choice of doctors: a look inside the blackest box of consumer behavior. Med Care Res Rev 2014 Oct;71(5 Suppl):38S-64S. [doi: 10.1177/1077558713496321] [Medline: 23999489]
11. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, MA: Houghton Mifflin; 2002.
12. Emmert M, Meier F, Heider AK, Dürr C, Sander U. What do patients say about their physicians? an analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
13. Jiménez FR, Mendoza NA. Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products. Journal of Interactive Marketing 2013 Aug;27(3):226-235. [doi: 10.1016/j.intmar.2013.04.004]
14. 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]
15. Ludwig S, de Ruyter K, Friedman M, Brüggen EC, Wetzels M, Pfann G. More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. J Mark 2013 Jan;77(1):87-103. [doi: 10.1509/jm.11.0560]
16. Dou X, Walden JA, Lee S, Lee JY. Does source matter? Examining source effects in online product reviews. Comput Human Behav 2012 Sep;28(5):1555-1563. [doi: 10.1016/j.chb.2012.03.015]
17. Kozinets RV, de Valck K, Wojnicki AC, Wilner SJ. Networked narratives: Understanding word-of-mouth marketing in online communities. J Mark 2010 Mar;74(2):71-89. [doi: 10.1509/jmkg.74.2.71]
18. Lee K, Koo D. Effects of attribute and valence of e-WOM on message adoption: Moderating roles of subjective knowledge and regulatory focus. Comput Human Behav 2012 Sep;28(5):1974-1984. [doi: 10.1016/j.chb.2012.05.018]
19. Sparks BA, Browning V. The impact of online reviews on hotel booking intentions and perception of trust. Tourism Management 2011 Dec;32(6):1310-1323. [doi: 10.1016/j.tourman.2010.12.011]
20. Zhang JQ, Craciun G, Shin D. When does electronic word-of-mouth matter? A study of consumer product reviews. J Bus Res 2010 Dec;63(12):1336-1341. [doi: 10.1016/j.jbusres.2009.12.011]
21. Cheung CM, Thadani DR. The impact of electronic word-of-mouth communication: A literature analysis and integrative model. Decis Support Syst 2012 Dec;54(1):461-470. [doi: 10.1016/j.dss.2012.06.008]
22. Dutta-Bergman M. Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. J Med Internet Res 2003 Sep;5(3):e21 [FREE Full text] [doi: 10.2196/jmir.5.3.e21] [Medline: 14517112]
23. Eastin MS. Credibility assessments of online health information: The effects of source expertise and knowledge of content. J Comput Mediat Commun 2001;6(4) [FREE Full text] [doi: 10.1111/j.1083-6101.2001.tb00126.x]
24. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
25. Wathen CN, Burkell J. Believe it or not: Factors influencing credibility on the Web. J Am Soc Inf Sci 2002;53(2):134-144. [doi: 10.1002/asi.10016]
26. King RA, Racherla P, Bush VD. What we know and don't know about online word-of-mouth: A review and synthesis of the literature. Journal of Interactive Marketing 2014 Aug;28(3):167-183. [doi: 10.1016/j.intmar.2014.02.001]
27. 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 Nov;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
28. Schindler RM, Bickart B. Perceived helpfulness of online consumer reviews: The role of message content and style. J Consumer Behav 2012 Mar 21;11(3):234-243. [doi: 10.1002/cb.1372]

29. Archak N, Ghose A, Ipeirotis PG. Deriving the pricing power of product features by mining consumer reviews. Manage Sci 2011 Aug;57(8):1485-1509. [doi: 10.1287/mnsc.1110.1370]
30. Detz A, López A, Sarkar U. Long-term doctor-patient relationships: patient perspective from online reviews. J Med Internet Res 2013 Jul;15(7):e131 [FREE Full text] [doi: 10.2196/jmir.2552] [Medline: 23819959]
31. Papathanassis A, Knolle F. Exploring the adoption and processing of online holiday reviews: A grounded theory approach. Tourism Management 2011 Apr;32(2):215-224. [doi: 10.1016/j.tourman.2009.12.005]
32. Lee E, Shin SY. When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo. Comput Human Behav 2014 Feb;31:356-366. [doi: 10.1016/j.chb.2013.10.050]
33. Lee J, Park D, Han I. The effect of negative online consumer reviews on product attitude: An information processing view. Electron Commer Res Appl 2008 Sep;7(3):341-352. [doi: 10.1016/j.elerap.2007.05.004]
34. Bickart B, Schindler RM. Internet forums as influential sources of consumer information. Journal of Interactive Marketing 2001 Jan;15(3):31-40. [doi: 10.1002/dir.1014]
35. Lockie M, Waiguny M, Grabner-Kräuter S. How style, information depth and textual characteristics influence the usefulness of general practitioners' reviews. Australasian Marketing Journal (AMJ) 2015 Feb;23(2). [doi: 10.1016/j.ausmj.2015.01.005]
36. Duan W, Gu B, Whinston AB. Do online reviews matter? - An empirical investigation of panel data. Decis Support Syst 2008 Nov;45(4):1007-1016. [doi: 10.1016/j.dss.2008.04.001]
37. Godes D, Mayzlin D. Using online conversations to study word-of-mouth communication. Marketing Science 2004 Nov;23(4):545-560. [doi: 10.1287/mksc.1040.0071]
38. Zhu F, Zhang X. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. J Mark 2010 Mar;74(2):133-148. [doi: 10.1509/jmkg.74.2.133]
39. Hsiao KL, Chuan ‐ Chuan Lin J, Wang XY, Lu H, Yu H. Antecedents and consequences of trust in online product recommendations. Online Information Review 2010 Nov 30;34(6):935-953. [doi: 10.1108/14684521011099414]
40. McFadden DL, Train KE. Consumers' evaluation of new products: learning from self and others. J Polit Econ 1996;104(4):683 [FREE Full text]
41. Lederman R, Fan H, Smith S, Chang S. Who can you trust? Credibility assessment in online health forums. Health Policy and Technology 2014 Mar;3(1):13-25. [doi: 10.1016/j.hlpt.2013.11.003]
42. Evans JS. Dual-processing accounts of reasoning, judgment, and social cognition. Annu Rev Psychol 2008;59:255-278. [doi: 10.1146/annurev.psych.59.103006.093629] [Medline: 18154502]
43. Jun SH, Vogt C. Travel information processing applying a dual-process model. Annals of Tourism Research 2013 Jan;40:191-212. [doi: 10.1016/j.annals.2012.09.001]
44. Klaczynski PA. Analytic and heuristic processing influences on adolescent reasoning and decision-making. Child Dev 2001;72(3):844-861. [Medline: 11405586]
45. Petty RE, Wegener DT, Fabrigar LR. Attitudes and attitude change. Annu Rev Psychol 1997;48:609-647. [doi: 10.1146/annurev.psych.48.1.609] [Medline: 9046570]
46. Eysenbach G, Powell J, Kuss O, Sa E. 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]
47. Flanagin AJ, Metzger MJ. Digital Media, Youth, and Credibility. Cambridge, MA: MIT Press; 2008. Digital media and youth: Unparalleled opportunity and unprecedented responsibility URL: http://www.comm.ucsb.edu/faculty/flanagin/CV/FlanaginMetzger2008%28DMYCch1%29.pdf [accessed 2015-03-25] [WebCite Cache ID 6XIh2QJFr]
48. Dutta-Bergman MJ. The impact of completeness and Web use motivation on the credibility of e-Health information. J Commun 2004 Jun;54(2):253-269. [doi: 10.1111/j.1460-2466.2004.tb02627.x]
49. Pornpitakpan C. The persuasiveness of cource credibility: A critical review of five decades' evidence. J Appl Soc Psychol 2004 Feb;34(2):243-281. [doi: 10.1111/j.1559-1816.2004.tb02547.x]
50. Xu Q. Should I trust him? The effects of reviewer profile characteristics on eWOM credibility. Comput Human Behav 2014 Apr;33:136-144. [doi: 10.1016/j.chb.2014.01.027]
51. Slater MD, Rouner D. How message evaluation and source attributes may influence credibility assessment and belief change. Journal Mass Commun Q 1996 Dec 01;73(4):974-991. [doi: 10.1177/107769909607300415]
52. Hovland CI, Weiss W. The influence of source credibility on communication effectiveness. Public Opin Q 1951;15(4):635-650. [doi: 10.1086/266350]
53. Eisingerich AB, Bell SJ. Maintaining customer relationships in high credence services. Journal of Services Marketing 2007 Jul 10;21(4):253-262. [doi: 10.1108/08876040710758559]
54. Morgan RM, Hunt SD. The commitment-trust theory of relationship marketing. J Mark 1994 Jul;58(3):20. [doi: 10.2307/1252308]
55. Hwang J, Yoon Y, Park N. Structural effects of cognitive and affective reponses to web advertisements, website and brand attitudes, and purchase intentions: The case of casual-dining restaurants. Int J Hosp Manag 2011 Dec;30(4):897-907. [doi: 10.1016/j.ijhm.2011.01.011]
56. Ohanian R. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. J Advert 1990 Oct;19(3):39-52. [doi: 10.1080/00913367.1990.10673191]

57. (Jimmy) Xie H, Miao L, Kuo P, Lee B. Consumers' responses to ambivalent online hotel reviews: The role of perceived source credibility and pre-decisional disposition. Int J Hosp Manag 2011 Mar;30(1):178-183. [doi: 10.1016/j.ijhm.2010.04.008]
58. Lai VS, Li H. Technology acceptance model for internet banking: an invariance analysis. Information & Management 2005 Jan;42(2):373-386. [doi: 10.1016/j.im.2004.01.007]
59. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, NY: Guilford Publications, Inc; 2013.
60. Zhao X, Lynch Jr J, Chen Q. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. J Consum Res 2010 Aug;37(2):197-206. [doi: 10.1086/651257]
61. Grabner-Kräuter S, Waiguny MK. "I believe more in factual reviews"– But not so much when the reviewer is similar to the reader and the product Is hedonic. In: Banks IB, De Pelsmacke P, Okazaki S, editors. Advances in Advertising Research. Volume V. Wiesbaden, Germany: Springer Fachmedien Wiesbaden; 2015:89-102.

Abbreviations

ANOVA: analysis of variance
PRW: physician-rating website
LLCI: lower level confidence interval
R: unstandardized regression coefficient
SE: standard error
ULCI: upper level confidence interval

Edited by G Eysenbach; submitted 29.10.14; peer-reviewed by M Emmert, S Agboola; comments to author 11.12.14; revised version received 21.01.15; accepted 04.02.15; published 09.04.15.

Please cite as:
Grabner-Kräuter S, Waiguny MKJ
Insights Into the Impact of Online Physician Reviews on Patients' Decision Making: Randomized Experiment
J Med Internet Res 2015;17(4):e93
URL: http://www.jmir.org/2015/4/e93/
doi:10.2196/jmir.3991
PMID:25862516

Original Paper

The eHealth Enhanced Chronic Care Model: A Theory Derivation Approach

Perry M Gee[1], RN, PhD; Deborah A Greenwood[2], RN, PhD; Debora A Paterniti[3], PhD; Deborah Ward[4], RN, FAAN, PhD; Lisa M Soederberg Miller[5], PhD

[1]School of Nursing, Division of Health Sciences, Idaho State University, Pocatello, ID, United States

[2]Diabetes Program Coordinator, Sutter Medical Foundation, Sacramento, CA, United States

[3]Center for Healthcare Policy and Research, UC Davis Health System, University of California, Davis, Sacramento, CA, United States

[4]Betty Irene Moore School of Nursing, UC Davis Health System, University of California, Davis, Sacramento, CA, United States

[5]Human Ecology, Human Development Unit, University of California, Davis, Davis, CA, United States

Corresponding Author:
Perry M Gee, RN, PhD
School of Nursing
Division of Health Sciences
Idaho State University
921 South 8th Avenue
Pocatello, ID, 83209
United States
Phone: 1 208 282 2182
Fax: 1 208 282 4476
Email: geeperr@isu.edu

Abstract

Background: Chronic illnesses are significant to individuals and costly to society. When systematically implemented, the well-established and tested Chronic Care Model (CCM) is shown to improve health outcomes for people with chronic conditions. Since the development of the original CCM, tremendous information management, communication, and technology advancements have been established. An opportunity exists to improve the time-honored CCM with clinically efficacious eHealth tools.

Objective: The first goal of this paper was to review research on eHealth tools that support self-management of chronic disease using the CCM. The second goal was to present a revised model, the eHealth Enhanced Chronic Care Model (eCCM), to show how eHealth tools can be used to increase efficiency of how patients manage their own chronic illnesses.

Methods: Using Theory Derivation processes, we identified a "parent theory", the Chronic Care Model, and conducted a thorough review of the literature using CINAHL, Medline, OVID, EMBASE PsychINFO, Science Direct, as well as government reports, industry reports, legislation using search terms "CCM or Chronic Care Model" AND "eHealth" or the specific identified components of eHealth. Additionally, "Chronic Illness Self-management support" AND "Technology" AND several identified eHealth tools were also used as search terms. We then used a review of the literature and specific components of the CCM to create the eCCM.

Results: We identified 260 papers at the intersection of technology, chronic disease self-management support, the CCM, and eHealth and organized a high-quality subset (n=95) using the components of CCM, self-management support, delivery system design, clinical decision support, and clinical information systems. In general, results showed that eHealth tools make important contributions to chronic care and the CCM but that the model requires modification in several key areas. Specifically, (1) eHealth education is critical for self-care, (2) eHealth support needs to be placed within the context of community and enhanced with the benefits of the eCommunity or virtual communities, and (3) a complete feedback loop is needed to assure productive technology-based interactions between the patient and provider.

Conclusions: The revised model, eCCM, offers insight into the role of eHealth tools in self-management support for people with chronic conditions. Additional research and testing of the eCCM are the logical next steps.

(J Med Internet Res 2015;17(4):e86) doi:10.2196/jmir.4067

KEYWORDS

chronic disease; mobile health; eHealth; telemedicine; personal health records; social networks; education of patients

Introduction

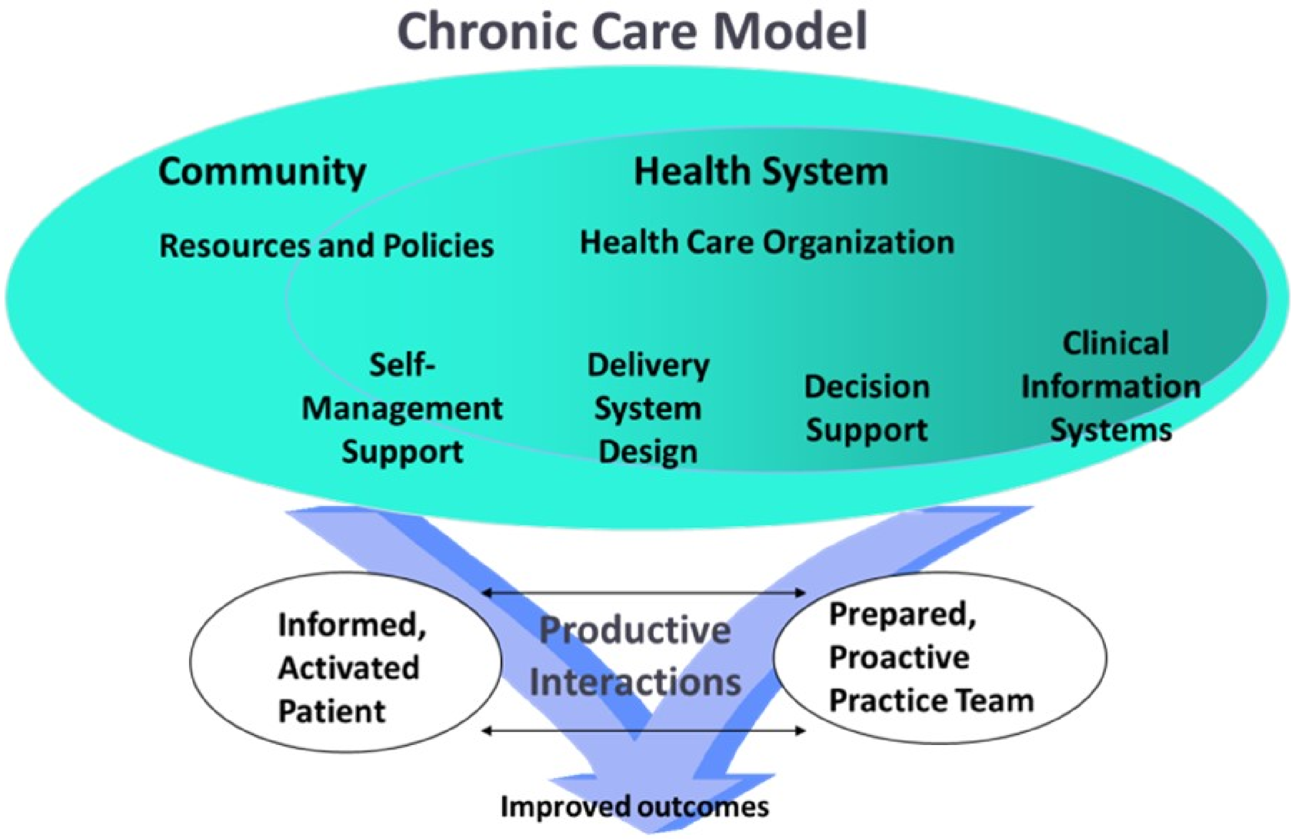
Background

Chronic illness is a burden on individuals and society; nearly half of Americans have at least one chronic illness accounting for more than three-fourths of America's health care spending [1,2]. The Chronic Care Model (CCM) is a well-established and validated framework that illustrates a comprehensive approach to caring for the chronically ill that supports increased functional and clinical outcomes. The model includes six key interdependent components: (1) community resources, (2) health system support, (3) self-management support, (4) delivery system design, (5) decision support, and (6) clinical information systems (Figure 1).

The CCM places chronic care in the context of the community where the person will receive health care services and with the health systems involved in that care. The CCM highlights the importance of "Self-Management Support"—giving patients the knowledge, confidence, and skills for self-management of their condition. "Delivery System Design" is also important to promote a patient-centered interdisciplinary team approach to care. "Decision Support" is needed to assure providers and patients have access to the most current and relevant evidenced-based guidelines for care, Finally, the model emphasizes the role of "Clinical Information Systems" to provide access to data, information, and knowledge needed to improve health. Effective and productive patient and provider interactions are the heart of the CCM and the key to improving outcomes [3].

Over the last decade, the CCM has been implemented and evaluated in a variety of settings in both domestic and international studies. The CCM has proven to be a useful framework for patient empowerment, self-management support, and improving clinical and behavioral outcomes [4-9]. The purpose of this review is to update the CCM with emerging eHealth technologies. This goal is consistent with the self-management support for chronic disease using technology tools suggested in both the Affordable Care Act (ACA) and the Health Information Technology for Economic and Clinical Health Act (HITECH Act) [10-12].

Figure 1. The Chronic Care Model. Developed by the MacColl Institute, ©ACP-JSIM Journals and Books, reprinted with permission from ACP-ASIM Journals and Books.



eHealth for Chronic Illness

Leaders and policy makers on a global scale are strongly encouraging the use of eHealth technologies. Australia, Europe, South Korea, and the United States all have strong eHealth initiatives that are developing policy for using information technology to improve health and health care systems. The Washington think tank "eHealth Initiative" (promoting policy focused on research and education in eHealth), the Institute of Medicine (IOM), and the Agency for Healthcare Research and Quality (AHRQ) also recommend the use of eHealth as a tool to support self-management in chronic illness [13-15]. Large systematic reviews conducted by the AHRQ determined that eHealth tools can improve patient engagement and health outcomes, however, more research is needed [14,15]. Jimison et al [14] also identified that eHealth technology interventions must contain a closed or complete feedback loop (CFL) to have an impact on chronic illness outcomes. A complete feedback loop contains five stages: (1) transmission of data and information regarding the health status of the consumer, (2) interpretation of data and information using previously established knowledge and/or wisdom and use of evidence-based standards, (3) address the specific need of the individual consumer, (4) timely feedback to the consumer addressing their requirements, and (5) regular repetition of the feedback loop [14].

Despite the strong push for eHealth, there is no standard definition, which hinders research and implementation. Comprehensive systematic reviews have identified as many as 51 different definitions for eHealth in the literature [13,16-18]. Overall, the literature describes the definition of eHealth to be very broad and encompassing, ranging from the very business-oriented to more clinically focused. The authors of this paper have developed a definition of eHealth for chronic illness self-management: To promote positive health outcomes by using a new frame of mind that incorporates information and communication technologies in the presence of a complete feedback loop and enables the use of data and information, to generate health management knowledge and wisdom.

Previous eHealth definitions by Eysenbach and Eng, and the Informatics language from Staggers and Thompson [19-21] influenced the definition. In the literature, the components of eHealth typically consist of use of the Internet, telemedicine, and communication [16]. In the IOM report *Health Literacy, eHealth, and Communication: Putting the Consumer First*, the round table members noted that the eHealth Initiative was guided by Wagner’s vision of the CCM and used as the “blueprint” for eHealth to support chronic illness [13]. For enhanced use with the CCM, the authors suggest specific components highlighted in the information technology and communication literature including use of the Internet for health information, social networking, telehealth, mHealth (including wearable devices), electronic health records (EHRs), and electronic personal health records/patient portals (PHRs).

Use of the Internet for Self-Management Support

The Internet serves as a conduit for self-management support, connecting providers and consumers to secure portals, health applications, social networks, and large databases. Roughly 80% of adults have sought health information on the Internet, including 62% of adults with a chronic illness; of those, 75% of the chronically ill surveyed stated their most recent Internet encounter affected decisions about the self-management of their condition [22-24]. The Internet is also the vehicle used by many adults for access to social networking sites.

Social Networking or eHealth Communities

Social networking or virtual communities are newer components of eHealth. A recent study reported in JAMA regarding the diabetes online community (DOC), “TuDiabetes”, found that the use of the social network augmented hypoglycemia surveillance among the members of that virtual community [25]. To date, most research on the impact of social networks has been descriptive but there are a few studies that have shown improved health outcomes; no studies to date have shown adverse effects on consumers/patients [26]. One randomized controlled trial (RCT) followed overweight and sedentary adults and found that the use of an online community helped maintain adherence to the program and that the participants had lower attrition from the study [27]. In recent descriptive studies, virtual community members with diabetes and heart disease found that the environment was useful for asking questions, reporting personal experiences, and even supported eHealth literacy [28,29]. Social networking may be an effective tool to encourage consumer empowerment and promote patient-centered care [30,31].

Telehealth

One well-researched component of eHealth is in the area of telehealth, sometimes called telemedicine, which has been used extensively as an intervention across many aspects of health care. Telehealth (telecommunication, videoconferencing, remote monitoring, etc) can range from performing a detailed physical examination either synchronously or asynchronously, to using videoconferencing (audio/video technology) for the delivery of a class or training to individuals or groups in a remote setting [32,33].

The telehealth field is challenged with a clear definition. A recent study by Doarn et al [34] found there are seven United States government definitions for telemedicine. A standard definition of telehealth to facilitate the use and research eHealth tools is essential. To add strength to this movement, a federal telemedicine group was commissioned, FedTel, and legislation has recently been introduced to Congress to establish federal telehealth standards [34,35].

Telehealth has been especially effective in the management of diabetes. A recent systematic review evaluating 15 RCTs described that hemoglobin A1c (A1C), a laboratory examination that measures average blood glucose over the past 2-3 months, improved when telehealth interventions incorporated more elements of structured self-monitoring of blood glucose [33]. Telehealth also lends itself to use by all members of the health care team. In a recent RCT, Tang et al [36] found that nurse-led, multi-disciplinary telehealth interventions were effective in improving A1C outcomes. In another nurse-led telehealth intervention designed for high-risk dialysis patients, the participants reported being more empowered and better able to provide needed self-management of illness [37]. In a recent RCT, Young et al found that a telehealth nurse coaching model for people with diabetes produced higher self-efficacy scores in the control group than for those who received the usual care [38]. Health care leaders, clinicians, and policymakers view telehealth as a powerful resource for improving health outcomes, health care quality, and to promote patient engagement [15,39,40].

Mobile Health

Mobile health (mHealth) components of eHealth span a broad spectrum of technologies. mHealth includes technology that is wireless, mobile [41], or wearable (eg, sensors, medication pumps, or wristbands that monitor physical activity). mHealth also includes thousands of health apps designed for mobile devices. The market for mHealth apps is anticipated to grow 25% per year with no foreseeable end in sight [42]. mHealth is a “disruptive innovation” providing entrée to Internet-based health resources to groups who previously had barriers to these tools; 60% of Americans gain access to the Internet using a laptop, tablet computer, or mobile phone [43]. Older adults, Hispanics, and African-Americans are adopting mobile technologies at a faster rate than the general population [43]. Progress in the area of mobile phone text messaging has created a surge in using the tool for health self-management. Several recent studies and systematic reviews have reported modest health outcome improvement using text messaging as a targeted intervention [44-46].

Electronic Health Records/Personal Health Records (EHR/PHR)

The EHR is an electronic longitudinal record of care and patient information that may be shared across multiple health care settings [47]. The tethered PHR, or patient portal, is a component of the EHR that communicates with the provider’s EHR or is integrated within the provider’s EHR and provides access to health records for patients/consumers and/or caregivers. The Markle Foundation [48] defined the PHR as “an electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized in a private, secure and confidential environment” (p. 14)(Markle Foundation, 2008)(Markle Foundation, 2008).

Health care leaders and policy makers highly recommend PHR use as a management tool for chronic illness [49-52]. However, the PHR literature including systematic reviews related to use and health outcomes were inconclusive and the few RTCs conducted on PHR interventions failed to make the case for efficacy [53]. While the scientific evidence does not demonstrate that using PHRs can improve health outcomes, the literature does support the impact on secondary outcomes related to self-management support [36]. In an observational cohort study of 8705 subjects, Sarkar et al [54] found that patients with diabetes who use the PHR to refill their prescriptions had better medication adherence. In spite of some demonstration for the efficacy of PHRs in optimizing care, the PHR literature, including systematic reviews related to use and health outcomes, has been inconclusive [53,55].

With the national and international push to implement eHealth technologies into the current care environment, there is an opportunity to augment the established CCM with integration of eHealth technology components. The remainder of this paper will describe how adding eHealth components to the CCM may provide more self-management tools for the person with chronic illness.

Methods

Theory Derivation

The Theory Derivation process was used to bring together the related eHealth concepts and to grasp the relatively new phenomenon of using eHealth tools for the self-management of chronic illness [56]. Theory derivation is a structured set of procedures where one chooses a parent theory or model that is used to guide the development of a new model or theory supported by a comprehensive understanding of the current literature [56]. In this paper, the CCM was carefully examined and supporting components were extrapolated for the development of a new model. Additionally, a methodical review of a wide range of literature was conducted. A draft framework was then developed and expanded by a continued review of new literature and evaluation of the established and new components of the revised model.

A thorough review of the published literature since 2000 was conducted using CINAHL, Medline, OVID, EMBASE, PsychINFO, Science Direct, and selected “grey” literature including government reports, industry reports, legislation, etc. The review involved using the search terms “CCM or Chronic Care Model” AND “eHealth” and then we searched the specific identified components of eHealth and Chronic disease self-management support (Virtual communities”, “Virtual health communities”, “e-Communities”, “on-line communities”, social networking”, “Telemedicine”, Telehealth”, “Internet use for health”, “mHealth”, “Electronic health records”, “Personal health records”, “Patient portals”, “User training”, “Technology”, “Chronic Illness”, “Chronic disease”, and “Self-management support”). Selection criteria included review papers, randomized controlled trials, cohort studies, cross-sectional studies, and qualitative studies. The researchers independently identified papers based on framework, design, sample, measures, and fit with self-management support and chronic illness. The CCM was carefully studied in the literature and then key components of the current CCM were used to provide a framework for new model construction.

Results

Summary

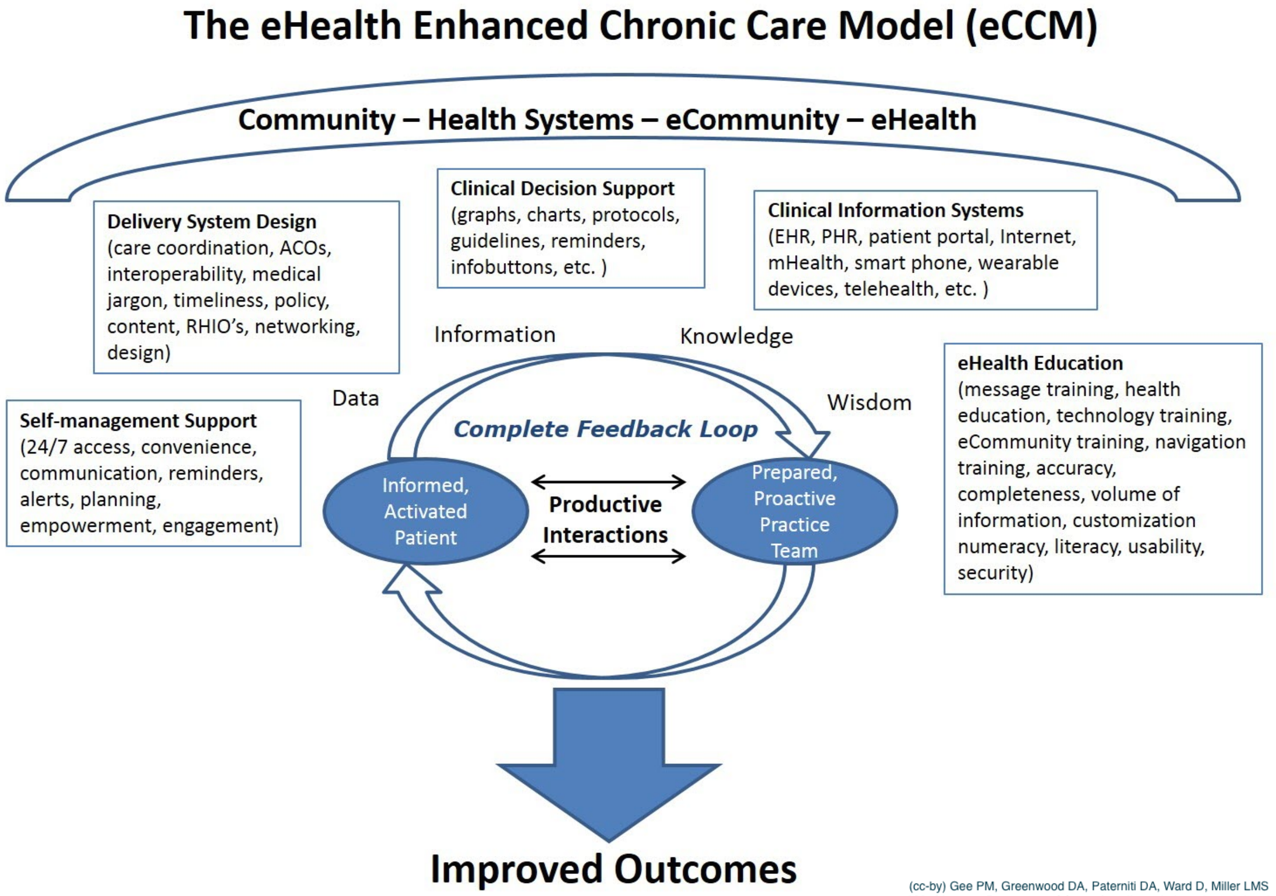
We identified 260 papers, but excluded 63.5% (165/260) due to concerns about study design, sample size, and/or methods. Overall, with the exception of telehealth interventions, there was a noted heterogeneity of methods and approaches used. We organized the literature into the components of the CCM, highlighting the role that eHealth tools and concepts can play in each component, modifying and adding components where needed to capture the emerging eHealth literature.

Adding the eCommunity and an Informatics Framework

The CCM has two major components: Community and Health Systems (Figure 1). The role of the community in the CCM is to provide support for patient engagement or activation and for self-management [57]. Based on the literature, the notion of community should be expanded to include online community and health-related social networks, or eCommunity (Figure 2). A total of 72% of American adults who use the Internet are already using social networks [58]. In a recent PEW poll, 26% of respondents stated they went online to observe the health postings about someone else's medical condition and 16% search the Internet to find another person with the same health ailment [59,60]. Hu, Bell, Kravitz and Orrange [61] found that in a survey of 505 participants in an online support group the adult members accessed the group to prepare for upcoming medical appointments.Virtual communities including "TuDiabetes" or "PatientsLikeMe" are already supporting thousands of chronically ill adults; both groups have vigorous research activities [25,62]. Health care technology leaders and policy makers are touting a new PHR 2.0 with expected growth of 221% over the next 3 years [63]. The concept of PHR 2.0 will contain the typical components of the current PHR/patient portal systems but add social networking, gaming, and e-visits [63].

Chronically ill adults who are activated, educated, engaged, and empowered, or e-patients, are already using eHealth tools [64,65]. E-patients, together with their providers, community, and social networks, have the ability to generate a collective knowledge and wisdom about their health care self-management needs greater than any of them working alone [65,66]. Figure 2, the eHealth enhanced CCM, contains the terms data, information, knowledge, and wisdom (DIKW). These words illustrate that the DIKW framework is underpinning the process of data and information used to create new knowledge and ultimately the collective wisdom to improve health outcomes [65,66].

Figure 2. The eHealth Enhanced Chronic Care Model. Created by Gee, P M; Greenwood, D A; Paterniti, D A; Ward, D; and Miller, L M S (JMIR, 2015). Adapted from The Chronic Care Model (see Figure 1).



(cc-by) Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LMS
J Med Internet Res 2015;17(4):e86, http://www.jmir.org/2015/4/e86/

Health Systems Enhancements

In the literature, it was noted that the health system must be designed to support organizations and providers and enable them to be prepared and proactive and to foster productive interactions with consumers; this is essential to improve health outcomes [67]. Kaiser Permanente and the Veterans Health Administration have made strategic efforts to implement eHealth technologies (PHRs, mHealth, telehealth, and Internet use) to improve access to care, reduce costs, and empower patients [68-71]. Between 2003 and 2010, Kaiser Permanente designed a system where patients could better manage their own health using a PHR [69,72]. Research at Kaiser showed a 25.3% reduction in the number of face-to-face office visits in primary care due to PHR use [73]. These findings suggest that health systems who consciously implement and encourage the use of eHealth technologies may achieve a higher level of patient engagement, satisfaction, and self-management support.

Delivery System Design Enhancements

"Meaningful Use" requires redesign of health care delivery systems to meet the emerging eHealth technologies to be implemented nation-wide. Meaningful Use Requirements state that providers are eligible for up to a US $44,000 payment from the centers of Medicare and Medicaid Services to implement provider electronic health record systems [10,74]. Stage II Meaningful Use further requires that providers and health care organizations implement a patient/provider communication portal system by 2014.

Access to and control over personal health data is a theme described by some health care consumers. This is an instance where the literature identified policy change and interoperability—the exchanging of data and information electronically between health care systems—as needed to provide the consumer more autonomy over their health data and information [75-77]. Access and control over personal health data are concepts similar to the “environmental factors” described by Tang and colleagues in their influential foundational PHR article [50]. Interoperability, lack of resources at the provider level, and PHR design and policy issues are factors outside the control of the individual participant and currently under the control of the provider or health care system. First, the decision makers in health systems or delivery systems should create policies to facilitate the correction of incorrect or missing data in the PHR. These policies would empower patients to work with their providers to assure the patient record is correct and current. Some health care organizations have had positive experiences with open access for consumers to the entire EHR/PHR including the provider notes [78]. Again, this is a policy issue that may promote productive interactions, engagement, and mutual trust.

Health system leaders can improve access issues for consumers by promoting policies that will encourage providers to release results sooner. Meaningful Use policy is currently encouraging more implementation of EHR/PHRs. Health care leaders are working on integration and interoperability of these records across health care systems and among individual providers [79].

Self-Management Support Enhancements

A review of the literature helped the authors to identify the core ideas of patient engagement and health self-management, empowered individuals, and the tools and knowledge to impact their own health. Research on PHRs shows improved patient engagement essential for self-management support [80]. The PHR encouraged engagement by facilitating preparation for appointments, tracking of laboratory results and diagnostic studies, encouraging involvement in preventive care and screening, and encouraging consumers to suggest a course of treatment with their providers [81,82].

The informed, “activated” patient is a key component of the CCM. Patient activation is the level of skills, knowledge, and confidence a person has in managing one’s own chronic illness [83]. The highly activated patient therefore is engaged, informed, and confident in their ability to self-manage their own condition [84]. Hibbard, Stockard, Mahoney, Tusler [85] developed a Patient Activation Measure (PAM) to determine the levels of patient activation. The use of a PHR can increase patient activation. An RCT of patients assigned a PHR as an intervention resulted in higher PAM scores compared to the control group, especially with those who started with lower scores [86]. Additional eHealth components such as telehealth and mHealth applications may also be useful in self-management support and to promote patient engagement [33,87]. The findings suggest that the use of a PHR can promote an informed, activated patient and augment the CCM in the areas of self-management support and productive interactions.

Clinical Decision Support Enhancements

Originally, the CCM identified clinical decision support (CDS) as a method to assure providers had access to the most current evidence-based clinical guidelines, protocols, and standard of care [67]. A study of the literature suggests the eCCM component of CDS should incorporate patient/consumer specific needs as follows: (1) visual access to data, (2) access to protocols, (3) care standards and evidence for self-management, (4) info buttons that access clinical guidelines, and (5) reminders for both the patients and providers. An RCT by Holbrook et al [88] found that people with type 2 diabetes had better outcomes when their intervention included a Web-based clinical decision support system shared by the patient and provider. Fox [22,23] found that chronically ill adults frequently go online to health sites to help make decisions about self-management of their condition. The Institute of Medicine [13,89,90] in three separate reports recommends incorporating eHealth tools for the promotion of CDS for both patients and providers to improve safety and self-management support.

Clinical Information Systems Enhancements

Originally, the clinical information systems (CIS) element of the CCM primarily focused on registries, databases, and systems in place to support the access to protocols and current standards of care. With rapid expansion of the eHealth components of the EHR/PHR, partially due to Meaningful Use implementations, the opportunity exists to engage with and evaluate these tools as part of the CCM. In 2001, 18% of provider offices had implemented EHRs; that number is up to 78% as of 2013 and about half of those implementations are meeting the Meaningful Use criteria [91]. Tethered patient portals/PHRs are part of the Meaningful Use stage II requirements and are dependent on EHR implementations to be a useful part of the enhanced CCM [50]. Other eHealth components such as telehealth and mobile devices are also on the rise. The inclusion of such tools in the CCM as part of the CIS element is a logical next step and one that has been proven to improve health outcomes [14]. Like the growth of PHRs, mHealth apps and mobile phones are expected to grow at a rate of 47.6% over the next six years [92].

Addition of eHealth Education to the CCM

Based on findings in the literature, an additional suggested major enhancement to the CCM is the addition of the support element “eHealth Education” (Figure 2). With health systems offering eHealth tools and consumers seeking eHealth solutions, providing the chronically ill adult with eHealth skills is needed [13]. Health literacy is essential for eHealth. Low health literacy is a long-standing problem in the current health care system [13]; 90 million Americans have poor health literacy—trouble understanding and managing their own health [93]. In a systematic review of eHealth interventions, Jacobs et al (2014) found that it is feasible to use eHealth tools to improve overall health literacy. Older adults make up the vast majority of those with chronic illness. And, while older adults are increasing their use of the Internet, social networking, and mobile phones, there is a gap in the literature in the evaluation of eHealth literacy for the older adult [94]. Choi and DiNitto studied older adults who are home-bound or poor and found opportunities for providing equipment and training that may improve eHealth literacy [95].

One component of health literacy that is problematic for many is health numeracy—people's ability to understand numbers and mathematical principles in the management of their health care [96]. Findings by Lipkus, Samsa, and Rimer [97] also noted that highly educated individuals had numeracy problems. To compound this issue, eHealth tools give patients even more access to data, information, and knowledge that may be confusing. eHealth is changing so rapidly that researchers are recommending we re-assess how we should measure eHealth literacy including its numeracy component [98].

Training for both consumers and providers in how to construct and send Web-based and text messages that promote productive interactions may prove useful. Training on the selection and use of health-related websites may also prove useful. Adults are already using the Web to look for health information on the Internet and were not informed on how to identify who is providing the information and how to assess the quality of the information [99,100]. With the rapid expansion of mHealth apps for mobile phones, training on how to choose apps that can promote health outcomes may be needed for both patients and providers [101,102].

Training in eHealth is proven to increase confidence and self-efficacy in using the tools but training lags behind in the roll-out of new technologies to the general public [15,103]. In fact in telehealth applications, the lack of training was reported frequently as a barrier to use [15]. In a mixed-methods study, low-income patients with human immunodeficiency virus (HIV) who were trained to use the PHR had better self-efficacy, patient activation, and disease knowledge at follow-up [104]. To promote productive interactions between the informed, activated patient and the prepared, proactive practice team, the authors of this paper would also recommend provider team training in the use of eHealth tools. In a study involving randomly assigned group practices, the providers trained to use the eHealth tools had measurable changes in the effectiveness of their information management skills [105]. Training for both consumers and providers may improve the efficacy of the use of eHealth tools and should be considered in future research.

Communication and the Addition of the Complete Feedback Loop (CFL)

Findings in the literature show that interventions that include the complete feedback loop are required for technology to promote improved health status in the chronically ill [14]. A key factor in the CCM is productive interactions between patients and providers. The activated patient is best suited to participate in the cooperative effort. In a qualitative study among chronically ill patients with a variety of levels of patient activation, it was found that being in control and working in a cooperative partnership with the provider was consistent with those patients who had the highest PAM scores [84]. This finding is similar with Yellowlees' [106] definition of "mutual participation" where patients work in an equal, trusting, and cooperative manner using the Internet to facilitate communication (p. 117). The secure patient-provider email messaging portal in a PHR is an area where this mutual participation and cooperation will occur. A cross-sectional study found that patients who were working cooperatively with their providers and setting very specific and concrete goals and focusing on self-monitoring had much higher PAM scores [107]. The PHR, mobile devices, and text messaging are eHealth tools that can give consumers control over the timing and content of their messages.

Provider response times to messaging are very important to patients [108,109] and can negatively affect CFL communication cycle. Patient satisfaction with provider response times to patient messages and requests have been studied. One study noted that patient satisfaction using a patient portal email system is positively affected by shorter message response times [108]. Reti et al found that response times varied across health care organizations and that usual patient portal email messaging response times varied from 24 to 72 hours [110]. With the importance of the CFL for self-management support and productive interactions, we recommend enhancing the CCM with the formal insertion of the CFL into the model. Perhaps the visual representation of the CFL in the model surrounding the productive interactions will remind researchers and developers of eHealth tools they need to include this element into their interventions and research (see Figure 2).

Discussion

Principal Findings

The purpose of this paper was to use Theory Derivation process to review the chronic care and eHealth literature and to articulate how the CCM could be expanded to include eHealth tools. The research is clear in showing that eHealth technologies related to a variety chronic conditions can be used to enhance self-management and revise the CCM [87,111]. The evidence also suggests that eHealth tools can support productive patient-provider interactions and improve health outcomes [112,113].

This review and model development highlights several gaps in the literature. First, clear definitions for eHealth, telehealth, and PHRs are needed to move forward in formulating appropriate research questions. Second, a gap in the literature exists in the efficacy of using online health communities for self-management support. While social support itself is shown to improve engagement and health outcomes with adults who have chronic illness [114,115], little is known about whether social support offered in online health communities has the same effect. A review of the literature and research in this area are needed. Third, research findings where elements of the CCM were used in conjunction with the CFL need to be identified and evaluated. Jimison et al (2008) identified that eHealth interventions that included the CFL improved outcomes [14]. Greenwood, Young, and Quinn found that telehealth interventions for people with diabetes is the kind of eHealth intervention that can foster a CFL [33]. The CFL as it relates to eHealth interventions is an important component to assure the success of eHealth interventions and will require future research [14]. Last, health education and technology experts are needed to develop a curriculum to train patients/consumers to use the eHealth tools that have been shown to improve health outcomes for the chronically ill person. Additionally, health care providers will need training on how to implement eHealth interventions and how to educate their chronically ill patients to use these tools.

Limitations

Limitations of this Theory Derivation process started with the fact that an exhaustive literature review was not completed for each of the eHealth components or the new elements added to the eCCM. Additionally, new and important literature is being added daily and with the scope of this project being so large, focused attention to the new knowledge was difficult to track. The opportunity exists for researchers to now concentrate on systematic reviews of the literature and conduct research specifically focusing on the individual components of the new eCCM model.

Conclusion

In conclusion, there is strong evidence demonstrating that eHealth tools can strengthen and enhance the already successful CCM. Research to explicitly test the new eCCM and its components is the logical next step.

Acknowledgments

The authors would like to thank Dr Robert Bell for his assistance in providing knowledge and literature supporting patient-provider communication. We would also like to thank the Gordon and Betty Moore Foundation for their generous support of the Betty Irene Moore School of Nursing at the University of California, Davis.

Authors' Contributions

PG conceptualized the project, reviewed the literature, constructed the initial eCCM, and prepared the manuscript drafts. DG offered exceptional knowledge of eHealth, the CCM, and self-management support for persons with chronic illness and assisted in project design. DP assisted with concept and model and development and understanding the needs of the patient. DW assisted with her vast nursing knowledge and wisdom, and her expertise in the area of health policy. LM offered her knowledge of the needs of the older adult, and human and cognitive factors related to the use of technology, health literacy, and with project design and drafting and revising parts of the manuscript.

Conflicts of Interest

None declared.

References

1. Ward BW, Schiller JS, Goodman RA. Multiple chronic conditions among US adults: A 2012 update. Preventing chronic disease 2014;11(130389):1-4. [doi: 10.5888/pcd11.130389]
2. Hoyert DL, Xi J. Deaths: Preliminary data for 2011. Hayattsville, MD: National Center for Health Statistics; 2012. URL: http://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_06.pdf [accessed 2015-03-24] [WebCite Cache ID 6XGp50tSQ]
3. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4 [FREE Full text] [Medline: 10345255]
4. Piatt GA, Orchard TJ, Emerson S, Simmons D, Songer TJ, Brooks MM, et al. Translating the chronic care model into the community: results from a randomized controlled trial of a multifaceted diabetes care intervention. Diabetes Care 2006 Apr;29(4):811-817. [Medline: 16567820]
5. Coleman K, Austin BT, Brach C, Wagner EH. Evidence on the Chronic Care Model in the new millennium. Health Aff (Millwood) 2009;28(1):75-85 [FREE Full text] [doi: 10.1377/hlthaff.28.1.75] [Medline: 19124857]
6. Tsai AC, Morton SC, Mangione CM, Keeler EB. A meta-analysis of interventions to improve care for chronic illnesses. Am J Manag Care 2005 Aug;11(8):478-488 [FREE Full text] [Medline: 16095434]
7. Siminerio LM, Piatt G, Zgibor JC. Implementing the chronic care model for improvements in diabetes care and education in a rural primary care practice. Diabetes Educ 2005;31(2):225-234. [doi: 10.1177/0145721705275325] [Medline: 15797851]
8. Siminerio LM, Piatt GA, Emerson S, Ruppert K, Saul M, Solano F, et al. Deploying the chronic care model to implement and sustain diabetes self-management training programs. Diabetes Educ 2006;32(2):253-260. [doi: 10.1177/0145721706287156] [Medline: 16554429]
9. Stellefson M, Dipnarine K, Stopka C. The chronic care model and diabetes management in US primary care settings: a systematic review. Prev Chronic Dis 2013;10:E26 [FREE Full text] [doi: 10.5888/pcd10.120180] [Medline: 23428085]
10. Blumenthal D. Implementation of the federal health information technology initiative. N Engl J Med 2011 Dec 22;365(25):2426-2431. [doi: 10.1056/NEJMsr1112158] [Medline: 22187990]
11. Blumenthal D, Tavenner M. The meaningful use regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
12. Patient Protection and Affordable Care Act, 42 U.S.C. § 18001 et seq. 2010. URL: http://www.ihs.gov/ihcia/documents/home/PL111_148_PPACA.pdf [accessed 2015-03-25] [WebCite Cache ID 6XHw6tp8I]

13. Institute of Medicine. Health literacy, eHealth, and communication: Putting the consumer first: Workshop summary. Washington, DC: The National Academies Press; 2009.
14. Jimison H, Gorman P, Woods S. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved: Evidence report/technology assessment No. Rockville, MD: Agency for Healthcare Research and Quality; 2008:09-E004.
15. Finkelstein J, Knight A, Marinopoulos S, Gibbons MC, Berger Z, Aboumater H, et al. Enabling patient-centered care through health information technology: Evidence report/technology assessment No. Rockville, MD: Agency for Healthcare Research and Quality; 2012:12-E005.
16. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
17. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
18. Showell C, Nohr C. How should we define eHealth, does the definition matter? 2012 Presented at: Quality of Life thrugh Quality of Information; August 26-29, 2012; Pisa, IT.
19. Staggers N, Thompson CB. The evolution of definitions for nursing informatics: a critical analysis and revised definition. J Am Med Inform Assoc 2002;9(3):255-261 [FREE Full text] [Medline: 11971886]
20. 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]
21. Eng TR. Population health technologies: emerging innovations for the health of the public. Am J Prev Med 2004 Apr;26(3):237-242. [doi: 10.1016/j.amepre.2003.12.004] [Medline: 15026105]
22. Fox S. Online health search 2006. Washington, DC: Pew Internet and American Life Project; 2006. URL: http://www.pewinternet.org/files/old-media/Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf [accessed 2015-03-24] [WebCite Cache ID 6XGrogS9c]
23. Fox S. E-patients with a disability or chronic disease. Washington, DC: Pew Internet and American Life Project; 2007. URL: http://www.pewinternet.org/2007/10/08/e-patients-with-a-disability-or-chronic-disease/ [accessed 2015-03-24] [WebCite Cache ID 6XGrvpX0v]
24. Fox S, Purchell K. Chronic disease and the Internet. Washington, DC: Pew Internet and American Life Project; 2010. URL: http://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/ [accessed 2015-03-24] [WebCite Cache ID 6XGs4TSDE]
25. Weitzman ER, Kelemen S, Quinn M, Eggleston EM, Mandl KD. Participatory surveillance of hypoglycemia and harms in an online social network. JAMA Intern Med 2013 Mar 11;173(5):345-351. [doi: 10.1001/jamainternmed.2013.2512] [Medline: 23400234]
26. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. J Biomed Inform 2013 Dec;46(6):957-969 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.010] [Medline: 23702104]
27. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
28. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med 2011 Mar;26(3):287-292 [FREE Full text] [doi: 10.1007/s11606-010-1526-3] [Medline: 20945113]
29. Jorgensen CB, Hansen J, Spindler H, Andreasen JJ, Nielsen G, Dinesen B. Heart patients' experiences and use of social media in their rehabilitation: A qualitative study. 2013 Presented at: Scandinavian Conference on Health Informatics; August 20, 2013; Copenhagen, Denmark.
30. Lober WB, Flowers JL. Consumer empowerment in health care amid the internet and social media. Semin Oncol Nurs 2011 Aug;27(3):169-182. [doi: 10.1016/j.soncn.2011.04.002] [Medline: 21783008]
31. 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]
32. American Telemedicine Association. 2014. What is Telemedicine? URL: http://www.americantelemed.org/about-telemedicine/what-is-telemedicine [accessed 2015-03-24] [WebCite Cache ID 6XGsLsAVy]
33. Greenwood DA, Young HM, Quinn CC. Telehealth remote monitoring systematic review: structured self-monitoring of blood glucose and impact on A1C. J Diabetes Sci Technol 2014 Feb 21;8(2):378-389. [doi: 10.1177/1932296813519311] [Medline: 24876591]
34. Doarn CR, Pruitt S, Jacobs J, Harris Y, Bott DM, Riley W, et al. Federal efforts to define and advance telehealth--a work in progress. Telemed J E Health 2014 May;20(5):409-418. [doi: 10.1089/tmj.2013.0336] [Medline: 24502793]
35. Matsui D. Telehealth Modernization Act of 2013. 2013. URL: https://www.govtrack.us/congress/bills/113/hr3750/text [accessed 2015-03-24] [WebCite Cache ID 6XGt6DLDl]

36. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013 May 1;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]
37. Minatodani DE, Chao PJ, Berman SJ. Home telehealth: facilitators, barriers, and impact of nurse support among high-risk dialysis patients. Telemed J E Health 2013 Aug;19(8):573-578. [doi: 10.1089/tmj.2012.0201] [Medline: 23742629]
38. Young H, Miyamoto S, Ward D, Dharmar M, Tang-Feldman Y, Berglund L. Sustained effects of a nurse coaching intervention via telehealth to improve health behavior change in diabetes. Telemed J E Health 2014 Sep;20(9):828-834. [doi: 10.1089/tmj.2013.0326] [Medline: 25061688]
39. Computer Sciences Corporation. Telemedicine: An essential technology for reformed healthcare. Falls Church, VA: Computer Sciences Corporation; 2011. URL: http://www.gatelehealth.org/wp-content/uploads/2011/06/Telemedicine_An_Essential_Technology_for_Reformed_Healthcare.pdf [accessed 2015-03-24] [WebCite Cache ID 6XGtVpg3O]
40. Institute of Medicine. Fostering independence, participation, and health aging through technology: Workshop summary. Washington, DC: The National Academies Press; 2013.
41. US Department of Health and Human Services. Using health text messages to improve consumer health knowledge, behaviors, and outcomes: an environmental scan. 2014. URL: http://www.hrsa.gov/healthit/txt4tots/environmentalscan.pdf [accessed 2015-03-24] [WebCite Cache ID 6XGulFTrt]
42. Borel B. Health policy brief: mHealth and FDA guidance. 2013. URL: http://healthaffairs.org/healthpolicybriefs/brief_pdfs/healthpolicybrief_104.pdf [accessed 2015-03-24] [WebCite Cache ID 6XGv9h7LK]
43. Smith A. Mobile access 2010. Washington, DC: Pew Internet & American Life Project; 2010. URL: http://www.pewinternet.org/2010/07/07/mobile-access-2010/ [accessed 2015-03-24] [WebCite Cache ID 6XGw8yfVi]
44. de Jongh T, Gurol-Uranci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illness. Cochrane Database of Systematic Reviews 2012;12(CD007459):1-40. [doi: 10.1002/14651858.CD007459.pub2]
45. Siopis G, Chey T, Allman-Farinelli M. A systematic review and meta-analysis of interventions for weight management using text messaging. J Hum Nutr Diet 2015 Feb;28 Suppl 2:1-15. [doi: 10.1111/jhn.12207] [Medline: 24480032]
46. Nundy S, Razi RR, Dick JJ, Smith B, Mayo A, O'Connor A, et al. A text messaging intervention to improve heart failure self-management after hospital discharge in a largely African-American population: before-after study. J Med Internet Res 2013;15(3):e53 [FREE Full text] [doi: 10.2196/jmir.2317] [Medline: 23478028]
47. Healthcare Information and Management Systems Society. Electronic health records. 2014. URL: http://www.himss.org/library/ehr/?navItemNumber=13261 [accessed 2015-03-24] [WebCite Cache ID 6XGx5qCgn]
48. Markle Foundation. The personal health working group: Final report. 2003. URL: http://www.markle.org/publications/1429-personal-health-working-group-final-report [accessed 2015-03-24] [WebCite Cache ID 6XGxGe9hP]
49. Ball MJ, Smith C, Bakalar RS. Personal health records: empowering consumers. J Healthc Inf Manag 2007;21(1):76-86. [Medline: 17299929]
50. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
51. Kaelber DC, Jha AK, Johnston D, Middleton B, Bates DW. A research agenda for personal health records (PHRs). J Am Med Inform Assoc 2008;15(6):729-736 [FREE Full text] [doi: 10.1197/jamia.M2547] [Medline: 18756002]
52. Tang PC, Lansky D. The missing link: bridging the patient-provider health information gap. Health Aff (Millwood) 2005;24(5):1290-1295 [FREE Full text] [doi: 10.1377/hlthaff.24.5.1290] [Medline: 16162575]
53. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
54. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
55. Saparova D. Motivating, influencing, and persuading patients through personal health records: a scoping review. Perspect Health Inf Manag 2012;9:1f [FREE Full text] [Medline: 22783154]
56. Walker LO, Avant KC. Theory Derivation. Strategies for Theory Construction in Nursing. Boston, MA: Prentice Hall; 2011.
57. Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: translating evidence into action. Health Aff (Millwood) 2001;20(6):64-78 [FREE Full text] [Medline: 11816692]
58. Duggan M, Smith A. Social media update. Washington, DC: Pew Research Center; 2013. URL: http://www.pewinternet.org/2013/12/30/social-media-update-2013/ [accessed 2015-03-24] [WebCite Cache ID 6XGxtkTFu]
59. Fox S, Duggan M. Health online 2013. Washington, DC: Pew Internet & American Life Project; 2013. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2015-03-24] [WebCite Cache ID 6XGy1NtoN]

60. Fox S. The social life of health information. Washington, DC: Pew Research Center; 2011. URL: http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2015-03-24] [WebCite Cache ID 6XGy8RuHq]
61. Hu X, Bell RA, Kravitz RL, Orrange S. The prepared patient: information seeking of online support group members before their medical appointments. J Health Commun 2012;17(8):960-978. [doi: 10.1080/10810730.2011.650828] [Medline: 22574697]
62. Bove R, Secor E, Healy BC, Musallam A, Vaughan T, Glanz BI, et al. Evaluation of an online platform for multiple sclerosis research: patient description, validation of severity scale, and exploration of BMI effects on disease course. PLoS One 2013;8(3):e59707 [FREE Full text] [doi: 10.1371/journal.pone.0059707] [Medline: 23527256]
63. Market disruption imminent as hospitals and physicians aggressively adopt patient portal technology.: Frost & Sullivan; 2013. URL: http://www.frost.com/prod/servlet/press-release.pag?docid=285477570 [accessed 2015-03-24] [WebCite Cache ID 6XGyUChxK]
64. Ferguson T, Frydman G. The first generation of e-patients. BMJ 2004 May 15;328(7449):1148-1149 [FREE Full text] [doi: 10.1136/bmj.328.7449.1148] [Medline: 15142894]
65. Gee PM, Greenwood DA, Kim KK, Perez SL, Staggers N, DeVon HA. Exploration of the e-patient phenomenon in nursing informatics. Nurs Outlook 2012;60(4):e9-16. [doi: 10.1016/j.outlook.2011.11.005] [Medline: 22221955]
66. Matney S, Brewster PJ, Sward KA, Cloyes KG, Staggers N. Philosophical approaches to the nursing informatics data-information-knowledge-wisdom framework. ANS Adv Nurs Sci 2011;34(1):6-18. [doi: 10.1097/ANS.0b013e3182071813] [Medline: 21150551]
67. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, Part 2. JAMA 2002 Oct 16;288(15):1909-1914. [Medline: 12377092]
68. Pearl R. Kaiser Permanente Northern California: current experiences with internet, mobile, and video technologies. Health Aff (Millwood) 2014 Feb;33(2):251-257. [doi: 10.1377/hlthaff.2013.1005] [Medline: 24493768]
69. Christensen K, Silvestre A. Making health personal. In: Liang L, editor. Connected for health: Using electronic health records to transform care delivery. San Francisco, CA: Jossey-Bass; 2010:139-156.
70. Saranto K, Brennan PF, Casey E. Personal health information management: tools and strategies for citizens' engagement. 2009 Presented at: International Nursing Informatics Congress; June 26-July 1, 2009; Vanajanlinna, Finland.
71. Nazi KM, Woods SS. MyHealtheVet PHR: A description of users and patient portal use. In: AMIA Annual Symposium Proceedings. 2008 Presented at: AMIA Annual Symposium; 2008; Washington, DC.
72. Liang L. Kaiser Permanente Healthconnect: A primer. In: Connected for health: Using electronic health records to transform care delivery. San Francisco, CA: Jossey-Bass; 2010.
73. Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente Electronic Health Record: transforming and streamlining modalities of care. Health Aff (Millwood) 2009;28(2):323-333 [FREE Full text] [doi: 10.1377/hlthaff.28.2.323] [Medline: 19275987]
74. Blumenthal D, Tavenner M. The meaningful use regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
75. Patel VN, Abramson E, Edwards AM, Cheung MA, Dhopeshwarkar RV, Kaushal R. Consumer attitudes toward personal health records in a beacon community. Am J Manag Care 2011 Apr;17(4):e104-e120 [FREE Full text] [Medline: 21774099]
76. Patel VN, Dhopeshwarkar RV, Edwards A, Barrón Y, Sparenborg J, Kaushal R. Consumer support for health information exchange and personal health records: a regional health information organization survey. J Med Syst 2012 Jun;36(3):1043-1052. [doi: 10.1007/s10916-010-9566-0] [Medline: 20703633]
77. Brennan PF, Downs S, Casper G. Project HealthDesign: rethinking the power and potential of personal health records. J Biomed Inform 2010 Oct;43(5 Suppl):S3-S5 [FREE Full text] [doi: 10.1016/j.jbi.2010.09.001] [Medline: 20937482]
78. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
79. Office of the National Coordinator for Health Information Technology. Beacon community and EHR vendor collaboration: A catalyst for interoperability and exchange. 2012. URL: http://www.healthit.gov/sites/default/files/pdf/ehr-vendor-beacon-topic.pdf [accessed 2015-03-24] [WebCite Cache ID 6XH0IpJ5l]
80. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
81. Wagholikar A, Fung M, Nelson C. Improving self-care of patients with chronic disease using online personal health record. Australas Med J 2012;5(9):517-521 [FREE Full text] [doi: 10.4066/AMJ.2012.1358] [Medline: 23115588]
82. Komura K, Yamagishi A, Akizuki N, Kawagoe S, Kato M, Morita T, et al. Patient-perceived usefulness and practical obstacles of patient-held records for cancer patients in Japan: OPTIM study. Palliat Med 2013 Feb;27(2):179-184. [doi: 10.1177/0269216311431758] [Medline: 22179597]
83. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]

84. Dixon A, Hibbard J, Tusler M. How do people with different levels of activation self-manage their chronic conditions? Patient 2009 Dec 1;2(4):257-268. [doi: 10.2165/11313790-000000000-00000] [Medline: 22273246]
85. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
86. Solomon M, Wagner SL, Goes J. Effects of a Web-based intervention for adults with chronic conditions on patient activation: online randomized controlled trial. J Med Internet Res 2012;14(1):e32 [FREE Full text] [doi: 10.2196/jmir.1924] [Medline: 22353433]
87. Siminerio LM. The role of technology and the chronic care model. J Diabetes Sci Technol 2010 Mar;4(2):470-475 [FREE Full text] [Medline: 20307409]
88. Holbrook A, Thabane L, Keshavjee K, Dolovich L, Bernstein B, Chan D, et al. Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. CMAJ 2009 Jul 7;181(1-2):37-44 [FREE Full text] [doi: 10.1503/cmaj.081272] [Medline: 19581618]
89. Institute of Medicine. Key capibilities of an electronic health record system. Washington, DC: National Academies Press; 2003.
90. Institute of Medicine. Health IT and patient safety: Building safer systems for better care. Washington, DC: The National Academies Press; 2012.
91. Hsiao CJ, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001-2012. Hyattsville, MD: National Center for Health Statistics URL: http://www.cdc.gov/nchs/data/databriefs/db111.pdf [accessed 2015-03-24] [WebCite Cache ID 6XH0g67y0]
92. Grand View Research. mHealth market analysis and segment forecasts to 2020. 2014. URL: http://www.grandviewresearch.com/industry-analysis/mhealth-market [accessed 2015-03-24] [WebCite Cache ID 6XH0oiXZM]
93. Institute of Medicine. Health literacy: A prescription to end confusion. Washington, DC: National Academies Press; 2004.
94. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
95. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
96. Lipkus IM, Peters E. Understanding the role of numeracy in health: proposed theoretical framework and practical insights. Health Educ Behav 2009 Dec;36(6):1065-1081 [FREE Full text] [doi: 10.1177/1090198109341533] [Medline: 19834054]
97. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making 2001;21(1):37-44. [Medline: 11206945]
98. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
99. Rideout V, Neuman T, Kitchman M, Brodie M. e-Health and the elderly: How seniors use the internet for health information. Menlo Park, CA: Kaiser Family Foundation; 2005. URL: https://kaiserfamilyfoundation.files.wordpress.com/2013/01/e-health-and-the-elderly-how-seniors-use-the-internet-for-health-information-key-findings-from-a-national-survey-of-older-americans-survey-report.pdf [accessed 2015-03-24] [WebCite Cache ID 6XH19xCkF]
100. Miller LMS, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. J Aging Health 2012 Apr;24(3):525-541. [doi: 10.1177/0898264311428167] [Medline: 22187092]
101. Powell AC, Landman AB, Bates DW. In search of a few good apps. JAMA 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
102. Aitken M, Goauntlett C. Patient apps for improved healthcare: from novelty to mainstream. Parsippany, NJ: IMS Institute for Healthcare Informatics; 2013.
103. Chu A, Huber J, Mastel-Smith B, Cesario S. Partnering with Seniors for Better Health: computer use and Internet health information retrieval among older adults in a low socioeconomic community. J Med Libr Assoc 2009 Jan;97(1):12-20 [FREE Full text] [doi: 10.3163/1536-5050.97.1.003] [Medline: 19159002]
104. McInnes DK, Solomon J, Shimada SL, Petrakis BA, Bokhour BG, Asch SM, et al. Development and evaluation of an internet and personal health record training program for low-income patients with HIV or hepatitis C. Med Care 2013 Mar;51(3 Suppl 1):S62-S66. [doi: 10.1097/MLR.0b013e31827808bf] [Medline: 23407015]
105. Schifferdecker KE, Reed VA, Homa K. A training intervention to improve information management in primary care. Fam Med 2008 Jun;40(6):423-432 [FREE Full text] [Medline: 18773781]
106. Yellowlees PM. Your health in the information age - how you and your doctor can use the internet to work together. Bloomington, IN: iUniverse, Inc; 2008.
107. Parchman ML, Zeber JE, Palmer RF. Participatory decision making, patient activation, medication adherence, and intermediate clinical outcomes in type 2 diabetes: a STARNet study. Ann Fam Med 2010;8(5):410-417 [FREE Full text] [doi: 10.1370/afm.1161] [Medline: 20843882]
108. Liederman EM, Morefield CS. Web messaging: a new tool for patient-physician communication. J Am Med Inform Assoc 2003;10(3):260-270 [FREE Full text] [doi: 10.1197/jamia.M1259] [Medline: 12626378]

109. Liederman EM, Lee JC, Baquero VH, Seites PG. Patient-physician web messaging. The impact on message volume and satisfaction. J Gen Intern Med 2005 Jan;20(1):52-57 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40009.x] [Medline: 15693928]
110. Reti SR, Feldman HJ, Ross SE, Safran C. Improving personal health records for patient-centered care. J Am Med Inform Assoc 2010;17(2):192-195 [FREE Full text] [doi: 10.1136/jamia.2009.000927] [Medline: 20190063]
111. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
112. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013;15(2):e37 [FREE Full text] [doi: 10.2196/jmir.2281] [Medline: 23425685]
113. 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]
114. Nicklett EJ, Liang J. Diabetes-related support, regimen adherence, and health decline among older adults. J Gerontol B Psychol Sci Soc Sci 2010 May;65B(3):390-399 [FREE Full text] [doi: 10.1093/geronb/gbp050] [Medline: 19541672]
115. Strom JL, Egede LE. The impact of social support on outcomes in adult patients with type 2 diabetes: a systematic review. Curr Diab Rep 2012 Dec;12(6):769-781 [FREE Full text] [doi: 10.1007/s11892-012-0317-0] [Medline: 22949135]

Abbreviations

A1C: hemoglobin A1C
ACA: Affordable Care Act
AHRQ: Agency for Healthcare Research and Quality
CCM: Chronic Care Model
CDS: clinical decision support
CFL: complete feedback loop
CIS: clinical information system
DIKW: Data, Information, Knowledge and Wisdom Model
DOC: diabetes online community
eCCM: eHealth Enhanced Chronic Care Model
EHR: electronic health record
HITECH: Health Information Technology for Economic and Clinical Health Act
HIV: human immunodeficiency virus
IOM: Institute of Medicine
PAM: patient activation measure
PHR: electronic patient health record/patient portal
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 21.11.14; peer-reviewed by K Jiggins Colorafi, D Baker; comments to author 01.01.15; revised version received 21.01.15; accepted 07.02.15; published 01.04.15.

Please cite as:
Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LMS
The eHealth Enhanced Chronic Care Model: A Theory Derivation Approach
J Med Internet Res 2015;17(4):e86
URL: http://www.jmir.org/2015/4/e86/
doi:10.2196/jmir.4067
PMID:25842005

Corrigenda and Addenda

Figure Correction: 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

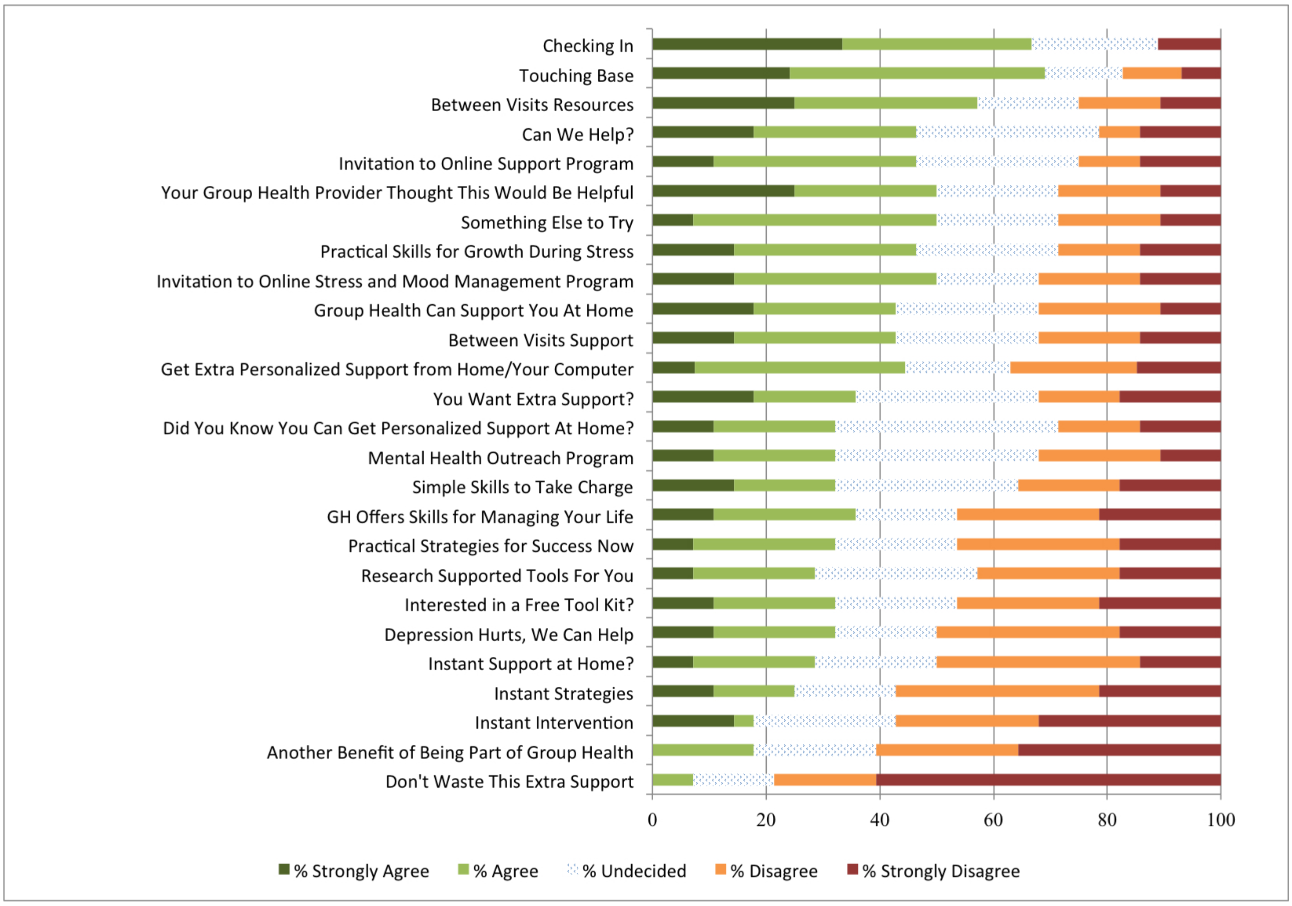
Related Article:

Correction of: http://www.jmir.org/2014/2/e42/

(J Med Internet Res 2015;17(4):e69) doi:10.2196/jmir.4412

The authors of "Designing Messaging to Engage Patients in an Online Suicide Prevention Intervention: Survey Results From Patients With Current Suicidal Ideation" (http://www.jmir.org/2014/2/e42/) have, during the final proofreading process, inadvertently added the same image file for Figures 1 and 2. Figure 1 has now been updated with the correct image, with the caption "I would like to open a message with Subject Line...". This error has been corrected in the online version of the paper on the JMIR website on April 13, 2015, together with publishing this correction notice. A correction notice has been sent to PubMed and the correct full-text has been resubmitted to Pubmed Central and other full-text repositories.

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



Edited by G Eysenbach; submitted 09.03.15; this is a non–peer-reviewed article;accepted 10.03.15; published 13.04.15.

Please cite as:
Whiteside U, Lungu A, Richards J, Simon GE, Clingan S, Siler J, Snyder L, Ludman E
Figure Correction: Designing Messaging to Engage Patients in an Online Suicide Prevention Intervention: Survey Results From Patients With Current Suicidal Ideation
J Med Internet Res 2015;17(4):e69
URL: http://www.jmir.org/2015/4/e69/
doi:10.2196/jmir.4412
PMID:25868120

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